University of Sussex

# Untangling the Physical Components of Galaxies using Infrared Spectra

**Peter Donald Hurley**

Submitted for the degree of Doctor of Philosophy

University of Sussex

February 2014

# Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

This thesis is written in the papers based style, as approved by the University.

Chapter two was published in Monthly Notices of the Royal Astronomical Society: *Principal Component Analysis and Radiative Transfer modelling of Spitzer IRS Spectra of Ultra Luminous Infrared Galaxies,* Hurley et al. 2012, Monthly Notices of the Royal Astronomical Society, 424, 2069-2078. I contributed to all aspects of this paper, with other authors giving comments, advice and providing access to data and models.

Chapter three was published in Monthly Notices of the Royal Astronomical Society: *Learning the Fundamental MIR Spectral Components of Galaxies with Non-Negative Matrix Factorisation*, Hurley et al. 2014, Monthly Notices of the Royal Astronomical Society, 437, 241-261. I contributed to all aspects of this paper, with other authors giving comments, advice and providing access to data and models.

Chapter four was published in Monthly Notices of the Royal Astronomical Society: *Herschel-SPIRE-Fourier Transform Spectroscopy of the nearby spiral galaxy IC342*, Rigopoulou et al. 2013, Monthly Notices of the Royal Astronomical Society, 434, 2051-2059. My contribution to this paper includes all of the modelling, interpretation as well as significant contributions to the text in Sections 4.3 and 4.4. I also made comments to the rest of the paper.

Chapter five was published in Nature: *A dust-obscured massive maximum-starburst galaxy at a redshift of 6.34*, Riechers et al. 2013, Nature, 496, 329-333. My contribution to this paper includes all the modelling associated with the $H_2O$ emission, joint interpret-

ation of the modelling results and I made contributions to the text. The most relevant work can be found in Section 5.2.1.

Signature:

Peter Donald Hurley

UNIVERSITY OF SUSSEX

Peter Donald Hurley, Doctor of Philosophy

# Untangling the Physical Components of Galaxies using Infrared Spectra

## Summary

The two main physical processes that underpin galaxy evolution are star formation and accretion of mass in active galactic nuclei (AGN). Understanding how contributions from these processes vary across cosmic time requires untangling their relative contributions.

The infrared part of the electromagnetic spectrum contains a number of AGN and star formation diagnostics e.g. emission lines from ionised gas or polyaromatic hydrocarbons (PAHs), and the shape of the continuum. Despite the higher resolution of data from Spitzer's IRS spectrograph, separating out emission from star formation and AGN is carried out using limited spectral features or simplistic templates. In the first part of this thesis, I show how sophisticated data analysis techniques can make full use of the wealth of spectral data.

I demonstrate how the popular multivariate technique, Principal Component Analysis (PCA), can classify different types of ultra luminous infrared galaxies (ULIRGs), whilst providing a simple set of spectral components that provide better fits than state-of-the art radiative transfer models. I show how an alternative multivariate technique, Non-Negative Matrix Factorisation (NMF) is more appropriate by applying it to over 700 extragalactic spectra from the CASSIS database and demonstrating its capability in producing spectral components that are physically intuitive, allowing the processes of star formation and AGN activity to be clearly untangled.

Finally, I show how rotational transition lines from carbon monoxide and water, observed by the Herschel Space Observatory, provides constraints on the physical conditions within galaxies. By coupling the radiative transfer code, RADEX, with the nested sampling routine, *Multinest*, I carry out Bayesian inference on the CO spectral line energy distribution ladder of the nearby starburst galaxy, IC342. I also show that water emission lines provide important constraints the conditions of the ISM of on one of the most distant starburst galaxies ever detected, HFLS3.

# Acknowledgements

There are many people to whom I owe a debt of gratitude. First and foremost I would like to thank my supervisor Seb Oliver. His enthusiasm and constant supply of ideas has always served as great motivation, especially in those dark moments of despair. I would also like to thank Duncan Farrah whose frank comments and advice I have come to rely on over the years. There are many others in astronomy who I would like to thank, including collaborators, friends and colleagues, but a special mention goes to Oskar Karczewski, Dimitra Rigopoulou, Kathy Romer, and Heather Campbell.

My curiosity and enjoyment of physics would not have developed without some inspirational teachers. A special thank you goes to my year eight science teacher, Geoff Ellis, whose experiments never failed to amuse or educate, and my G.C.S.E science teacher Dr Jon Sargent, whose enthusiastic lessons and stories started me on the long route towards a Ph.D.

My greatest thank you goes to my family and partner. To Charlotte Clarke, who has put up with my moaning and frantic writing during the final few months. My two brothers, Richard and Fraser, who never failed to bring me back to down to Earth when my head got too big, nor failed to cheer me up or entertain me when I needed it most. Finally, to my parents, whose unwavering support and encouragement has allowed me to pursue my career, despite some of life's great hurdles.

## Official acknowledgments

Funding in this thesis was provided by STFC studentship ST/F006977/1.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Prelude

Throughout our history, humans have tried to understand the cosmos. However, the greatest leap in our knowledge and understanding of the Universe, without doubt, occurred during the 20th Century.

At the end of the 19th Century, the Universe was believed to extend only as far as our own galaxy, was infinitely old, and would exist indefinitely. The Great Debate between Harlow Shapley and Heber Curtis in 1920 led to a dispute on whether the Universe extends to distances far greater than that occupied by the Milky Way, which was partially resolved with Edwin Hubble's discovery of Cepheid variable stars in the Andromeda Galaxy (M31). A further step change in our understanding came in 1929, when Hubble demonstrated the Universe was not static but expanding.

The notion of an expansion introduced the idea of an evolving Universe, which in turn led to the fundamental questions which underpin modern scientific cosmology: what is the origin of the Universe, and what is its eventual fate?

One hypothesis addressing the question of origin came in the form of Lemaître's 'Big Bang' theory (Lemaître, 1927). Lemaître proposed that the Universe began from a single point, from which it exploded out in a 'Big Bang' and has carried on expanding ever since. The hypothesis split the scientific community and it was not until the detection of the Cosmic Microwave Background (CMB) by Penzias & Wilson (1965) that the 'Big Bang' theory became generally accepted.

The CMB radiation is a relic of the early Universe from a time where matter and radiation were coupled together. At that time, baryonic matter consisted of ionised hydrogen and free electrons, which through continuous Thomson scattering, prevented photons from

propagating through space freely. As the Universe expanded, it cooled. Approximately 400,000 years after the Big Bang, it reached low enough temperatures for electrons to combine with hydrogen nuclei in a process termed recombination. After recombination, photons were able to travel unimpeded throughout the Universe. As a result of expansion, the wavelength of these photons has been stretched, redshifting the photons to microwave wavelengths, which we observe today as the CMB.

By the end of the 20th Century, the COsmic Background Explorer (COBE, Mather (1982); Gulkis et al. (1990)) satellite had measured the CMB to a high degree of accuracy and revealed that the CMB takes the form of a perfect blackbody spectrum and is almost entirely homogenous.

The beginning of the 21st Century saw the launch of the Wilkinson Microwave Anisotropy Probe (WMAP; Bennett et al. (2003)), and more recently the Planck observatory (Planck Collaboration et al., 2011) which has provided increasingly more precise observations of the CMB, refining the Big Bang theory and the current standard model of cosmology, referred to as ΛCDM.

The ΛCDM describes a Universe with an age of 13.7 Billion years and its total energy divided amongst 'dark energy' with as yet unknown nature, making up 69%, invisible or 'dark matter' constituting 26% and ordinary baryonic matter contributing only 5% (Planck Collaboration et al., 2013).

## 1.2 Galaxy Evolution

The CMB is a snapshot of the early Universe, taken before the formation of the complex structure of stars, galaxies and clusters of galaxies. The growth of structure began from Gaussian quantum fluctuations present in the early Universe, which we observe in the CMB as tiny temperature fluctuations. These fluctuations were traced out by dark matter, giving rise to areas of varying density. The over-densities of dark matter eventually grew through gravitational attraction to form large concentrations, called dark matter haloes.

As the Universe continued to expand and large scale structure continued to grow, baryonic matter began to accrete onto the haloes, where it cooled and collapsed further until it was dense enough to begin forming stars and galaxies.

The formation of galaxies created suitable conditions for complex physical and chemical processes such as star formation, supernovae, active galactic nuclei, magnetic fields etc complicating the gravitational growth of structure and giving us the rich, vibrant Universe we live in today.

Understanding how these physical processes impact the evolution of a galaxy is a large part of modern day extragalatic astronomy. This thesis concentrates on identifying and quantifying the processes of star formation and active galactic nuclei.

### 1.2.1 Star formation

Stars are the building blocks of galaxies and are both directly and indirectly involved in producing the heavy elements essential for life on Earth.

#### Molecular clouds

The formation of stars begins with a cloud of molecular gas and dust, consisting of predominantly molecular hydrogen, with the next most abundant molecule being carbon monoxide. The clouds either spontaneously collapse and form stars of its own accord, or the collapse can be triggered by dynamical mechanisms such as shock waves or ionisation fronts from supernovae, HII regions or planetary nebulae, a collision with another molecular cloud, outflows from the centre of galaxies or even the process of merging with another galaxy.

The process of gravitational collapse is accompanied by an increase of density and temperature of the gas in a pre-stellar cloud. Having reached higher densities, the cloud becomes far more efficient in molecular line emission cooling. Molecules present in the interstellar medium remove the excess heat through the process of collisional excitation, followed by radiative de-excitation. The energy levels of a molecule can have rotational, vibrational or electronic states. The rotational energy states require the least energy for excitation, making them the most relevant for colder clouds. Molecular hydrogen is symmetric and has no electric dipole moment which makes it an inefficient coolant at low temperatures (e.g. Kreckel et al., 2010).

Carbon monoxide (CO) does have an electric dipole moment and is therefore an efficient coolant for dense, cold clouds. As the CO molecules collide, the kinetic energy from the collisions excites the molecule into higher rotational energy states. The molecule than relaxes back down to less energetic states by emitting photons with wavelengths ranging from the submillimetre to millimetre. The strengths of the different CO rotational energy transitions produce a CO spectral line energy distribution (SLED) ladder. As CO is the best tracer for colder molecular clouds, the birth places of stars, the observations of the CO SLED has become important in our understanding of the process of star formation.

### Birth of stars

With the clouds beginning to cool through the process of molecular line emission, gravitational collapse will start to occur if the mass of the cloud exceeds the Jeans mass ($M_J$), given by:

$$M_J = 4 \cdot 10^5 (T^3/N)^{1/2} (M_\odot) \tag{1.1}$$

where $T$ is kinetic temperature and $N$ is number density.

The cloud will continue to contract and will fragment. The collapsing clouds are referred to as protostars. Through a complex process of cooling and gravitational collapse, the centre of the protostar eventually reaches a density and temperature high enough to undergo nuclear fusion turning the protostar into a star.

### Signatures of star formation

In order to understand how star formation has influenced galaxy evolution, we need to be able to measure the rate at which galaxies form stars, the star formation rate (SFR). Observing individual molecular clouds and individual stars is only possible in our own Galaxy. Observations in extragalactic astronomy are limited to the integrated effects of star formation over entire star formation regions, if not entire galaxies.

Different regions of the electromagnetic spectrum provide direct tracers of recent star formation. In the radio, thermal (free-free) radiation directly traces the ionising photon rate arising from newly formed massive stars and is largely extinction free (Condon, 1992). Another direct tracer is emission from the ultraviolet (UV), where emission is dominated by young stars (e.g. Kennicutt, 1998; Calzetti et al., 2005; Salim et al., 2007). However, the UV suffers from interstellar extinction, which varies considerably for different types of galaxy. The extinction is caused by absorption and reprocessing of stellar emission by material present in the interstellar medium. Most of the reprocessed stellar radiation absorbed by gas is emitted as optical and near IR emission lines from ionised regions surrounding the most massive stars. The H$\alpha$ recombination line is the main tracer of HII* regions, however other lines such as the [OII] forbidden line doublet at 372.7 nm are also used.

Beyond HII regions lie areas of the interstellar medium containing mostly neutral atomic gas which reprocesses the (6 eV $< h\nu <$13.6 eV) photons from massive stars that are not energetic enough to ionise hydrogen. These are the photon dominated regions (PDRs). As demonstrated in the schematic in Figure 1.1 the illuminated surface of a

*defined as areas where Hydrogen atoms are ionised

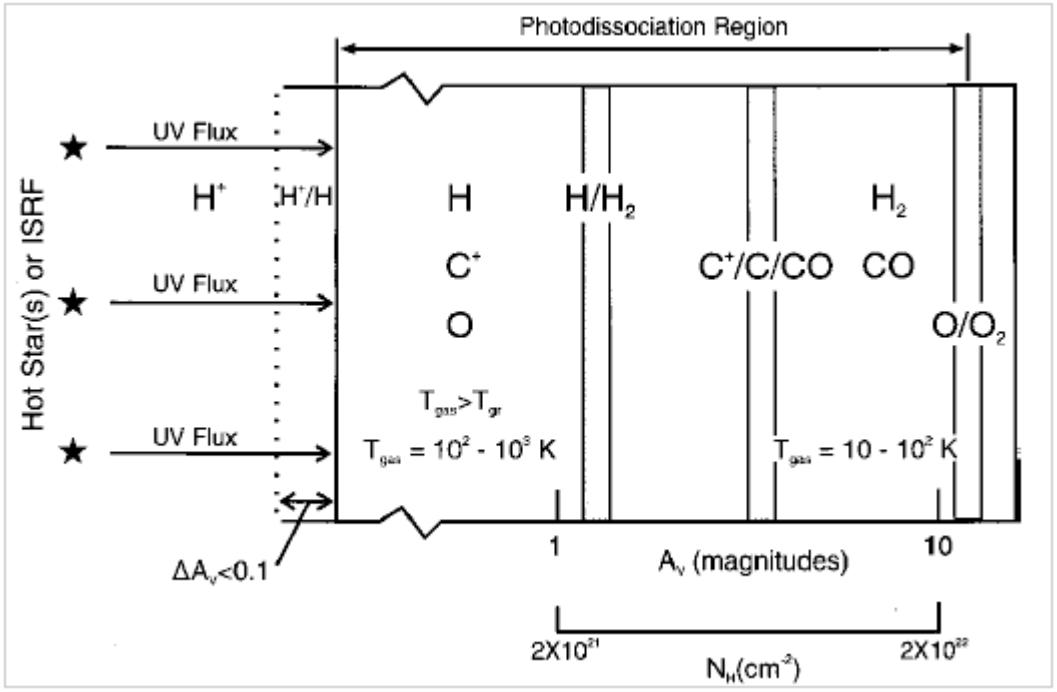


Figure 1.1: A schematic of a PDR, with the UV emission illuminating a cloud of gas from the left. The PDR transitions from atomic gas containing H, $C^+$ and O to molecules such as $H_2$ and CO. Figure taken from Tielens (2005).

PDR is made up of atomic gas such as atomic hydrogen, oxygen and ionised carbon. Further into the PDR, where most of the higher energy photons have been absorbed, molecular hydrogen and CO can begin to exist.

Most of the non-stellar infrared (IR) and the millimetre and submillimetre CO emission from galaxies originates from PDRs. The PDRs also produce fine structure lines such as [CI] 609, [CII] 158, and [OI] 63 $\mu$m; and the ro-vibrational lines of $H_2$ to name but a few. Theoretical models of PDRs such as Meijerink et al. (2007), model the various physical and chemical processes (e.g. heating through photo-electric emission by dust grains, fine structure line cooling, molecular line cooling and ion-molecule reactions, see Meijerink & Spaans (2005)) in order to predict the various line intensities associated with different parameters such as intensity of the illuminating source and metallicity. By matching the predicted models to observations of some of these lines, it can be ascertained as to whether the incident radiation is a result of on going star formation or whether other radiation fields such as X-rays from AGN are responsible for heating the interstellar medium (see section 1.2.2).

As well as gas, the interstellar medium also contains dust made up of a mixture of silicate grains and carbonaceous grains. These can range in size from very small grains,

consisting of tens to hundreds of atoms, to large grains with diameters exceeding one $\mu$m (Draine & Li, 2001; Li & Draine, 2001). The smaller grains with radii $< 0.025\mu$m contribute significantly to the extinction in the ultraviolet and are thought to be made up of mostly polycyclic aromatic hydrocarbon (PAH) like molecules. The PAHs are vibrationally excited following absorption of a single UV/visible photon and emit at distinct wavelengths in the mid-infrared, giving rise to what are generally referred to as the PAH features. The mid to far infrared wavelengths are dominated by thermal blackbody like continuum emission from the dominating silicate and carbonaceous grain populations.

#### Cosmic star formation history

We have seen that star formation is a fundamental process in galaxy evolution and that various indicators can help quantify the star formation rate of individual galaxies. These indicators can be used to trace the observational global star formation rate back with redshift, giving one of the fundamental relations relating star formation and galaxy evolution, the cosmic star formation rate density (CSFRD), $\dot{\rho_\star}(z)$. The quantity, which describes the average amount of mass that turns into stars per unit time per unit comoving volume at redshift $z$, has been estimated for redshifts ranging from 0 to 6.

Lilly et al. (1996) provided the first observational evidence of a varying CSFRD, followed shortly by Madau et al. (1996, 1998). Further evidence from combined observational evidence can also be found in Hopkins et al. (2006). The general consensus, still accepted today is that the CSFRD peaks at a redshift of around 2-3, with a value an order of magnitude greater than at present. At redshifts greater than 3, the CSFRD appears to decline, but there remains a high level of uncertainty above a redshift of 4.

In practice, the CSFRD is calculated from luminosity functions. Recent work by Burgarella et al. (2013) has shown the importance of using the luminosity functions from both the far ultraviolet and the infrared in order to get a complete view of the star formation history of the Universe. One complication in converting the infrared luminosity function into a corresponding star formation rate density is that it contains two components, the required star formation contribution and a contribution from one of the other fundamental galaxy processes, active galactic nuclei.

### 1.2.2 Active Galactic Nuclei

An important process that affects galaxy evolution is the accretion of mass onto super massive black holes (SMBH) at the centre of a galaxy. This is a highly energetic pro-

cess producing non-thermal radiation over the full wavelength range from X-ray to radio wavelengths and their feedback processes can have an important effect on the interstellar medium, thereby affecting other processes such as star formation as well as the general formation and evolution of a galaxy.

Galaxies which show signs of this type of activity are called active galaxies, while the small region specifically associated with the accretion process is referred to as the active galactic nucleus (AGN). Active galaxies form a small but important population of galaxies. The lifetime of AGN is believed to be short and many, if not all galaxies are thought to harbour a central SMBH, which may accrete mass at some point in their lifespan.

The population of AGN can be subdivided according into three main groups according to the observational properties of the galaxy, though many of these classifications have an historical background:

**Seyfert** galaxies, first investigated by Seyfert (1943), are spiral galaxies with bright nuclei. The spectral properties show strong emission lines associated with high excitation. The Seyfert class is divided into two distinct subclasses. Seyfert 1s have strong broad emission lines with a corresponding velocity on the order of 1000-5000kms$^{-1}$ for permitted lines such as hydrogen (Mo et al., 2010). For forbidden lines such as [OIII] 5007Å, widths are much narrower, with velocities on the order of a few hundred kms$^{-1}$. Seyfert 2s have permitted and forbidden lines which are both narrow, again on the order of a few hundred kms$^{-1}$ Khachikian & Weedman (1974). Other differences between the two types include stronger hard X-ray emission for type 1 Seyferts.

**Radio galaxies** are a class of active galaxies whose radio emission is relatively strong. Their optical spectra show similarities to Seyfert galaxies and are divided into broad-line radio galaxies and narrow-line radio galaxies. However the morphology of radio galaxies differs in that the host galaxies tends to be elliptical rather than spiral.

**Quasars** (Quasi-Stellar Radio source) are AGN objects with associated radio emission, emanating from jets from the central nucleus of an AGN. Historically, quasars were first discovered by Matthews & Sandage (1963), and were associated with faint, unresolved optical sources which were believed to be stellar like objects. Their optical spectra show similarities with Seyfert 1 like objects, however quasars are far more luminous, are blue and often variable, indicating that they are small.

**Unification scheme**

The present day consensus is that many of the different classifications of AGN are a consequence of an orientation effect (e.g. Rowan-Robinson, 1977). Figure 1.2 shows a schematic from Urry & Padovani (1995) showing the main features of an AGN. The supermassive black hole is the driving force behind an AGN, as first suggested by Salpeter (1964). The material falling into a SMBH forms an accretion disk due to the general angular momentum of the accreting matter. The material is heated and emits thermal radiation, but with a temperature profile dependent on radius within the disk. The resulting spectrum takes the form of a stretched blackbody, peaking in the optical or ultraviolet. Jets of material can be funnelled out of the accretion disc, possibly due to magnetohydrodynamic processes (Blandford & Payne, 1982; Camenzind, 1990), and emit radio and X-ray photons due to synchrotron and inverse Compton scattering, respectively. Clouds of gas lying above the accretion disk are subject to high velocities due to the gravitational potential of the SMBH, broadening the emission lines and giving rise to the broad line region. Further out, the velocity of the clouds of gas decreases and so the emission is narrower.

Further out from the accretion disk lies the obscuring torus, made of gas and dust. The material in the torus absorbs the continuum radiation from the accretion disk, and much of the broad line region. The reprocessing of the radiation by dust in the torus, gives rise to emission in the mid-infrared.

With the axisymmetric model of the AGN shown in Figure 1.2, Seyfert 1 like objects can be described by viewing the AGN face on, which gives a full view of the accretion disk, broad line regions and narrow line regions. Seyfert 2 like objects can be explained by viewing the AGN side-on, so that the dusty torus obscures the accretion disk and most of the broad line regions, while the narrow line regions, which are prevalent above the torus, can still be observed.

**Signatures of AGN**

We have discussed how AGN can produce emission at wavelengths ranging from X-ray to radio, but with observational features that vary depending on the type of AGN.

The X-ray emission produced in jets orthogonal to the accretion disk is associated with strong AGN activity (Brandt & Hasinger, 2005), and radio emission from the outflowing jets is another direct indicator of AGN activity. Spectral features such as the width of permitted hydrogen and forbidden [OIII] emission lines indicate Seyfert 1 and 2 like

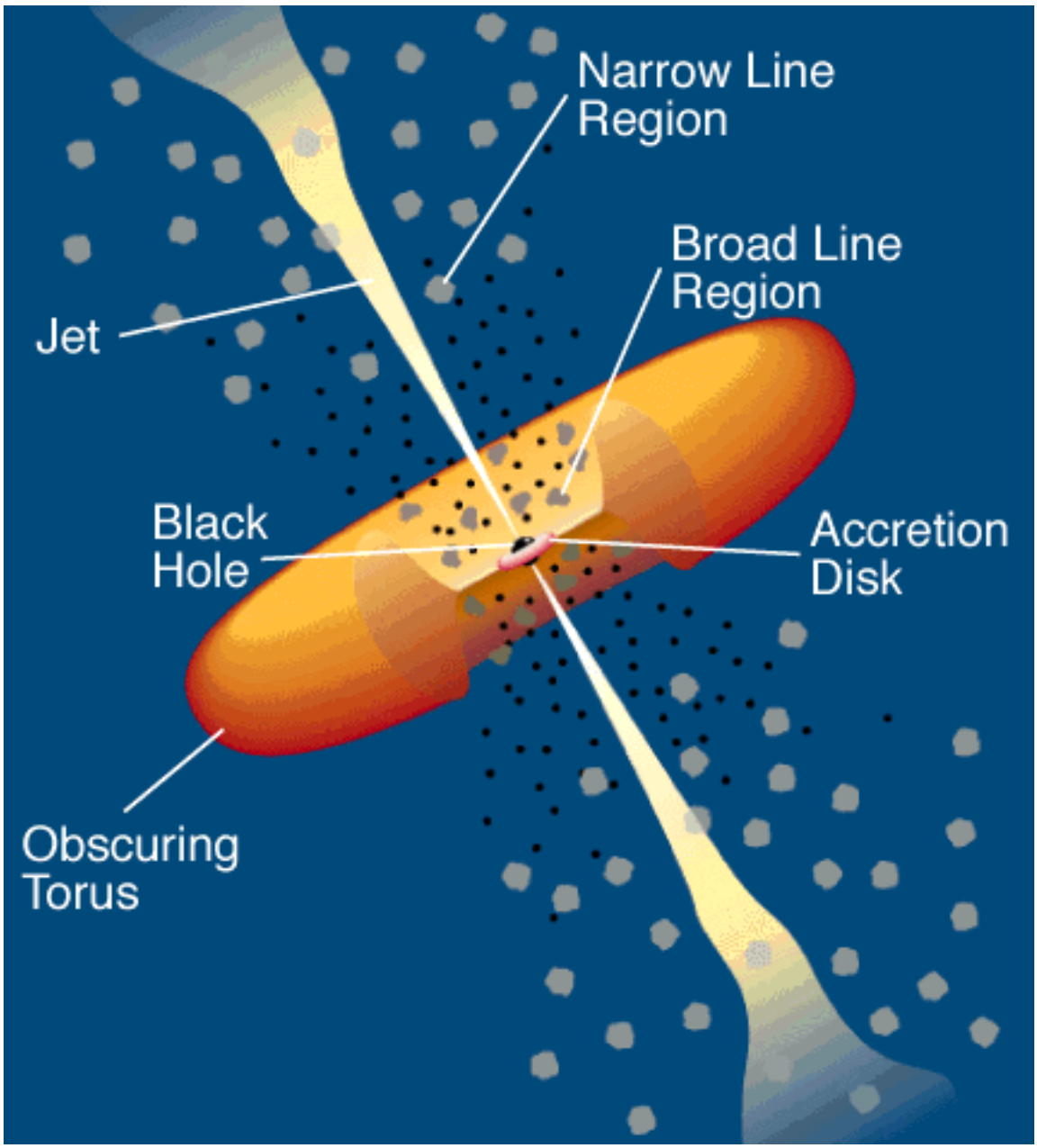


Figure 1.2: Schematic diagram for AGN. The SMBH is located at the centre, surrounded by an accretion disk. Slightly further out lie clouds of material, moving rapidly due to the gravitational potential of the black hole. These clouds give rise to broad line emission. Further out, lie clouds which move slower and give rise to narrow line emission. Jets of energetic material emit radio emission via the synchrotron process along the poles of the accretion disk, while further out and with the same orientation as the accretion disk lies an obscuring torus made up of dust and gas. Figure taken from Urry & Padovani (1995).

galaxies, while the presence of high-ionisation emission lines from the narrow-line regions such as [Ne V] and [OIV] can also act as tracers of AGN activity.

The shape of the continuum in the infrared is another signature. Radiative transfer models of the dusty torus (e.g. Efstathiou & Rowan-Robinson, 1995; Nenkova et al., 2002) have been successfully fitted to the optical to far-infrared SED of active galaxies.

Observing the far-infrared and sub-millimetre emission lines can also act as a tracer for AGN activity. Neutral gas in the interstellar medium can be exposed to X ray radiation from the AGN. The higher energy photons have a different chemical effect on the material than the lower energy photons associated with PDRs. Hard X-rays ($E > 1$ keV) from AGN can penetrate far deeper into gas clouds than UV photons, resulting in different thermal and chemical structures through the cloud compared to PDRs. The different structures result in different line intensities to PDRs and by comparing theoretical models of 'X-ray dominated regions' such as those in Meijerink & Spaans (2005), to observations of emission lines (including the CO SLED), it can be ascertained as to whether an AGN is heating the interstellar medium rather than photons from star formation.

### 1.2.3 The AGN-star formation connection

The processes of star formation and AGN can have drastic effects on the evolution of the host galaxy. It has been established that the two processes are linked (e.g. Kauffmann et al., 2003; Fernandes et al., 2001) through processes such as AGN feedback (i.e. the interaction between the interstellar medium and energy and radiation produced from the SMBH, see Fabian (2012) for a review).

Furthermore, there exists a population of galaxies which appear to be a composite of AGN activity and extreme star formation. These objects are often dusty and therefore ultra luminous (i.e. $L_{8-1000\mu m} > 10^{12} L_{\odot}$) in the infrared, hence known as ultra luminous infrared galaxies (ULIRGs). These objects are rare in the local Universe, but the associated luminosity function shows strong, positive evolution with redshift (e.g. Sanders, 1999), resulting in several hundred ULIRGs per square degree at $z > 1$ (Rowan-Robinson et al., 1997; Barger et al., 1998; Hughes et al., 1998; Eales et al., 2000; Fox et al., 2002; Floc'h et al., 2005). The increase in number density with redshift and their associated high SFRs means that ULIRGs make a significant contribution to the cosmic star formation rate density.

Untangling the contribution from AGN and star formation in individual galaxies is important if one is to understand the connections between the two. For dusty galaxies

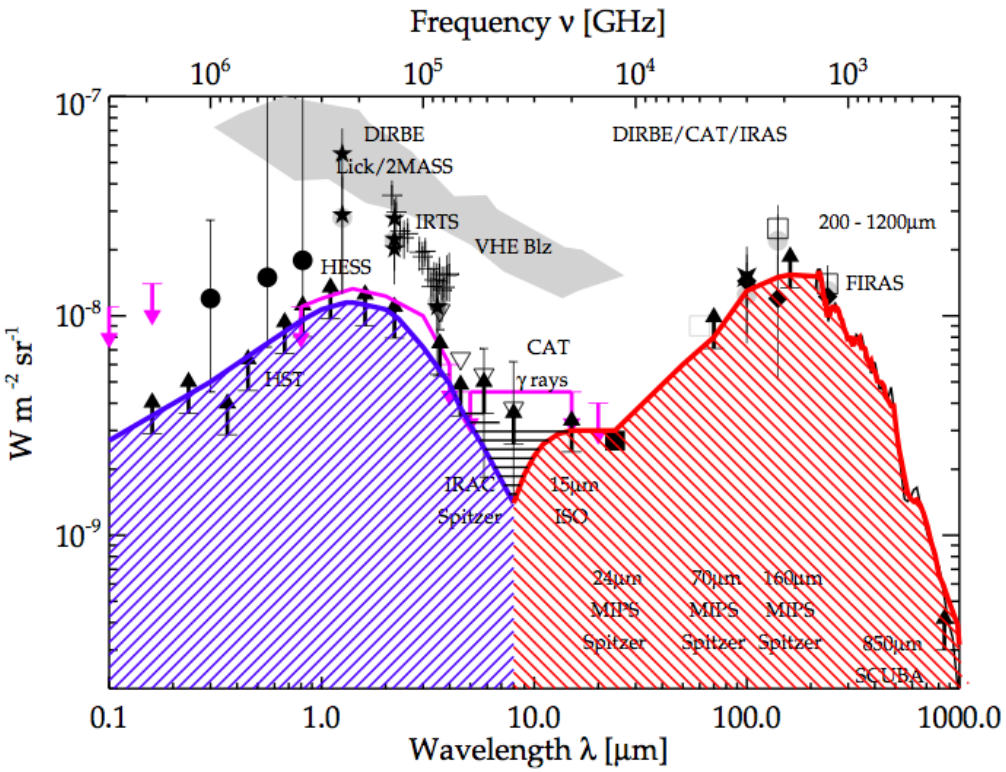


Figure 1.3: The cosmic optical and infrared background, showing the energy density as a function of wavelength. Taken from Dole et al. (2006).

such as ULIRGs, the AGN is heavily obscured by the dusty torus which blocks any X-ray or UV emission from the central region, while re-radiating the energy in the mid-infrared. With dust emission from star forming regions also contributing to the mid to far infrared wavelengths, untangling how much emission is related to each process becomes a challenging problem. Attempts to separate out the contributions to the mid-infrared luminosity function has already been carried out by Valiante et al. (2009) and Wu et al. (2011), through the use of fairly crude spectral templates based on individual galaxies.

## 1.3 Infrared Astronomy

To fully understand galaxy evolution, observations of galaxies across the entire electromagnetic spectrum are required. If one looks at the energy density of the extragalactic background light, it becomes apparent that the cosmic infrared background (CIB) contributes roughy the same as the cosmic optical background, as shown in Figure 1.3.

The Earth's atmosphere radiates and absorbs in mid and far infrared wavelengths, therefore prohibiting ground based observations for most parts of the infrared spectrum.

Studies at these wavelengths are therefore mostly restricted to the more expensive space based facilities. The following section details the space based based observatories covering the mid-far infrared wavelengths (e.g. $8 - 1000\mu$m).

The first infrared space telescope was the InfraRed Astronomical Satellite (IRAS; Neugebauer et al. (1984)) launched in 1983. IRAS observed the whole sky in four bands centred at 12, 25, 60 and 100 $\mu m$ and revealed new populations of galaxies which were optically faint but luminous in the infrared (Soifer et al., 1984).

The next major space based facility for infrared astronomy was the Infrared Space Observatory (ISO; Kessler et al. (1996)). Unlike IRAS, ISO provided spectroscopy for the mid-infrared part of the spectrum. The next major advancement came with the launch of the Spitzer Space Telescope (Werner et al., 2004). Spitzer provided near and mid-infrared photometry to greater depths than ever before, whereas its onboard Infrared Spectrograph (IRS; Houck et al. (2004)) provided higher resolution and more sensitive mid-infrared spectroscopy for thousands of galaxies.

Extragalactic spectra from ISO and IRS contain a wealth of spectral features, including the PAH features, high ionisation fine structure lines such as [NeV] and [OIV], broad absorption features associated with silicates, and underlying continua with a variety of shapes. Combinations of these different features have been used as diagnostics for characterising the power source behind the ULIRGs (Genzel et al., 1998; Rigopoulou et al., 1999; Farrah et al., 2007; Spoon et al., 2007; Farrah et al., 2008, 2009; Petric et al., 2011).

Other smaller space based facilities have included AKARI (Murakami et al., 2007), which surveyed the sky in six wavelength bands from mid to far-infrared and the Wide-field Infrared Survey Explorer (WISE, Wright et al. (2010)) which surveyed the sky at near to mid infrared wavelengths. However, the most recent advancement in infrared astronomy has been made with the ESA Herschel Space Observatory (Pilbratt et al., 2010), launched in 2009. Photometry from the PACS and SPIRE instruments has given an unprecedented view of the far-infrared Universe. Spectroscopy has also been possible with the SPIRE Fourier-Transform Spectrometer (FTS), covering 194 to 671 $\mu m$. This spectral region contains a number of high-J CO line transitions, allowing the peak of the CO SLED to be probed for the first time, which in turn enables modelling of the physical properties of the molecular ISM.

## 1.4 Statistical analysis and modelling

The ideology behind the work in this thesis is to use advanced statistical and modelling techniques to help untangle the contributions made from star formation and AGN activity.

The statistical techniques can be thought of as multivariate analysis techniques. By applying these techniques to astronomical data, one can make the most of the data at hand. The following section provides a general overview of some of the techniques used and their context within this thesis

In terms of modelling, the work in this thesis uses Bayesian inference. The Bayesian approach is more logical in terms of dealing with belief and uncertainty in model fitting and selection, and therefore provides a more robust indication on how well models are constrained by the data. From a physical perspective, knowing the uncertainty on parameters can be just as important as the knowing the values that give the best fit.

As the Bayesian approach underpins the modelling work in this thesis, this section also introduces the basics of Bayesian inference, and how it can be carried out on CO and $H_2O$ SLEDs with nested sampling. Finally, I discuss the principles of radiative transfer that underpin the models used for modelling the SLEDs.

### 1.4.1 Multivariate Statistics

As previously discussed, the IRS spectrograph on Spitzer has provided the astronomical community with observations of thousands of galaxies. The resulting datasets are of high dimensionality, with the number of dimensions being equal to the number of wavelength points. Most analyses of extragalactic IRS spectra have focused on using specific spectral features such as the PAH emission lines, while ignoring other dimensions.

Ideally, all dimensions should be taken into account in order to extract all the information contained in the data. Simultaneous analysis of high dimensional, or multivariate data is the guiding principle behind multivariate statistics. This area of statistics encompasses a wide range of techniques which take into account the relationships between the variables of a multi-dimensional dataset. These multivariate techniques can be used to uncover latent structure in datasets and reduce a large number of variables down to a smaller number of factors.

One of the most popular multivariate analysis techniques is Principal Component Analysis.

**Principal Component Analysis**

Although a dataset maybe highly dimensional, the data may lie on an almost planar subspace of lower dimensionality. Principal component analysis (PCA) finds the basis vectors which describe this linear subspace. Each datapoint can then be described in terms of these new basis vectors, or principal components. One of the main assumptions made by PCA is that the distribution of data points can be described by a multi-dimensional Gaussian. The basis vectors are then orthogonal components which describe this multi-dimensional Gaussian. The first component captures the direction of greatest variance, the second component captures the next most varying direction, whilst remaining orthogonal to the first. There are as many components as there are dimensions, however only a few 'principal' components may be needed to adequately describe the data, while the rest are considered 'noise'.

Algebraically, PCA of a dataset, $\{\mathbf{x}^n, n = 1, .., N\}$ with $N$ data points and $D$ dimensions, can be determined by calculating the eigenvectors and eigenvalues of the $D \times D$ covariance matrix of the dataset. The principal components are simply the eigenvectors, sorted in descending order by their corresponding eigenvalues. Choosing $M$ principal components results in an $M$ dimensional approximation of the dataset such that for datapoint $\mathbf{x}^n$:

$$\mathbf{x}^n \approx \mathbf{m} + \sum_{j=1}^{M} y_j^n \mathbf{b}^j \tag{1.2}$$

where $\mathbf{m}$ is the mean $D$ dimensional vector of the the dataset, $y_j^n$ is the weight for the $n$th object and $j$th component, and $\mathbf{b}^j$ is the $j$th principal component.

Having applied PCA to $\{\mathbf{x}^n, n = 1, .., N\}$, every object in the dataset is approximated by a linear combination of the $M$ principal components and can now be described by $M$ weights. By examining the principal components, one can ascertain what variables or dimensions are correlated, and are responsible for producing the greatest variation between objects in the dataset.

PCA has been used in astronomy for spectral classification of optical galaxies (e.g. Connolly et al., 1995; Bromley et al., 1998; Taghizadeh-Popp et al., 2012), while the first application to mid-infrared spectroscopy was carried out by Wang et al. (2011).

The first paper in this thesis, extends the work of Wang et al. (2011) by examining in more detail the number of components needed to adequately describe a sample of local ultra luminous infrared galaxies. We also fit the spectra with a combination of AGN dusty torus radiative transfer models described in Efstathiou & Rowan-Robinson (1995) and the Siebenmorgen & Krügel (2007) starburst radiative transfer models. We compare the

reconstructions from principal components to the best fits provided by radiative transfer models, and use the models to try and give a physical interpretation to the principal components. Finally, we utilise the reduction in dimensions to define a classification scheme for the IRS spectra of ULIRGs by using other sophisticated classification tools such as Gaussian Mixtures modelling (discussed in papers one and two) to learn the areas of principal component (PC) space associated with different, optically classified galaxies.

Having examined the intricacies of PCA, and its application to the IRS spectra of ULIRGs, it became clear that some of the assumptions required by PCA were not well suited for mid-infrared spectra. In particular, the principal components were not physically intuitive, which limits their use to data compression. Ideally, we would want to derive a set of components that have a clear physical interpretation. This was the motivation behind the second paper in this thesis, where we examined what other multivariate analysis techniques would be more suited to the mid infrared spectra of galaxies.

### Matrix Factorisation and Non-Negative Matrix Factorisation

Matrix factorisation techniques are a subset of multivariate analysis techniques. They seek to approximate a data matrix such that $\mathbf{X} \approx \mathbf{WH}$, or:

$$\mathrm{X}_{i\mu} \approx (\mathbf{WH})_{i\mu} = \sum_{a=1}^{r} \mathrm{W}_{ia}\mathrm{H}_{a\mu} \tag{1.3}$$

.

Where, $i$ is object index, $\mu$ is wavelength index and $a$ is component index. The matrix $\mathbf{H}$ can be thought of as a set of $r$ components that represent latent structure explicit in the dataset, and $\mathbf{W}$ are a set of weighting coefficients. Each object in the dataset can now be approximated by a linear combination of the derived components, $\mathbf{H}$.

PCA, although not strictly a matrix factorisation technique, can be viewed in this setup by substituting the data matrix $\mathbf{X}$ with the mean subtracted data matrix.

Non-negative matrix factorisation (NMF) attempts to solve the matrix factorisation equation 1.3, with the assumption that data, weights and components are all positive. This simple constraint can have drastic affects on the derived weights and components as reconstruction of $\mathbf{X}$ is forced to be purely additive, unlike PCA which allows both addition and subtraction.

This was demonstrated in Lee & Seung (1999) who showed that by applying NMF to a database of facial images, the NMF derived components resembled parts of the face, e.g. eye, nose mouth etc, unlike PCA whose components had holistic representations with no obvious physical interpretation.

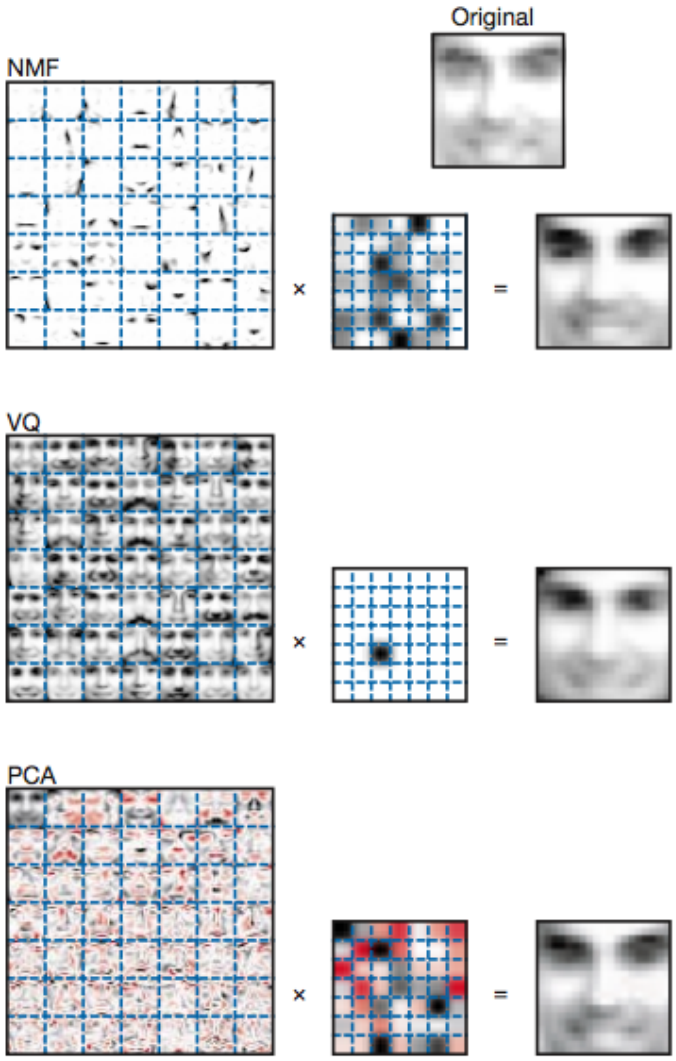


Figure 1.4: 49 basis images for NMF, vector quantization (VQ, a cluster based matrix factorisation where a face is forced to be made up of only one of the basis images) and PCA and learnt from 2,429 facial images, each consisting of 19 x 19 pixels. The basis images, seen on the left, are multiplied by a combinations of weights in order to reconstruct a particular face. NMF learns parts based basis images which resemble parts of the face unlike VQ or PCA. Positive values are illustrated with black pixels and negative values with red pixels. Figure taken from Lee & Seung (1999)

Mathematically, the goal of NMF is to minimise a cost function. The most widely used is the squared approximation error described in Lee & Seung (2001):

$$\chi^2 = \sum_{i\mu} \left( \mathrm{X}_{i\mu} - \sum_a \mathrm{W}_{ia}\mathrm{H}_{a\mu} \right)^2 \tag{1.4}$$

Minimising equation 1.4 requires a numerical technique to find local minima. Lee & Seung (2001) presented 'multiplicative update rules' for H and W. Upon each iteration, the rules are used to update H and W by a multiplicative factor to minimise equation 1.4.

The first application of NMF in astronomy was carried out by Blanton & Roweis (2007) who updated the popular NMF multiplicative algorithm from Lee & Seung (2001) to include uncertainties and allow heterogeneous datasets (e.g. optical spectra and photometric observations of galaxies at different redshifts) as an input. They also restricted the space of possible spectra to those predicted from high resolution stellar population synthesis models, thereby making the components of **H** an additive mixture of models.

In the second paper of this thesis, NMF is applied to the IRS spectra of 729 galaxies, with redshifts ranging from 0.01 to 0.2. Unlike Blanton & Roweis (2007), we do not restrict the subspace of possible components, and instead use NMF to blindly learn the components from the galaxy spectra. By doing so, we are using NMF as a blind source separation algorithm, where a galaxy spectrum is an additive mixture of components. The positive constraints is a more realistic model for the emission processes in the mid-infrared than the forced linear addition and subtraction required by PCA. As a result, the derived components are physically intuitive.

### 1.4.2 Bayesian Inference

Papers two, three and four make use of Bayesian inference. The following subsection outlines some of the important details. Bayesian inference can be split into two categories: parameter estimation and model selection both of which are commonplace in astronomy and cosmology and used throughout work in this thesis. The basis of Bayesian inference comes from one of the fundamental relations in probability theory, Bayes rule:

$$Pr(\Theta|D,H) = \frac{Pr(D|\Theta,H)Pr(\Theta|H)}{Pr(D|H)} \tag{1.5}$$

Where:

- $Pr(\Theta|D,H)$ is the posterior probability distribution of parameters $\Theta$ in model $H$, given data $D$

- $Pr(D|\Theta, H)$ is the likelihood distribution of $D$, given $\Theta$
- $Pr(\Theta|H)$ is the prior belief of $\Theta$
- $Pr(D|H)$ is a normalising factor, or Bayesian evidence (can be thought as 'how much $D$ increases our belief in $H$')

In parameter estimation, the goal is to obtain $Pr(\Theta|D, H)$ by sampling the unnormalised posterior with routines such as MCMC whilst ignoring the Bayesian evidence as it has no dependence on $\Theta$. In model selection, the Bayesian evidence is the desired quantity and is the integral of the likelihood $L(\Theta)$ over all parameter space defined by the prior. By averaging the likelihood over the parameter space, the Bayesian evidence automatically implements Occam's razor, i.e. a simpler theory with compact parameter space will have a larger evidence than a more complicated one, unless the latter is significantly better at explaining the data.

Calculating the Bayesian evidence is computationally challenging since it is a multi-dimensional integral, with the number of dimensions equal to the number of parameters. Although modified MCMC methods have been used to calculate the evidence, the number of samples required make the calculation at least an order of magnitude more costly than parameter estimation, while the behaviour of the likelihood distribution can also cause issues with the calculation (Skilling, 2004).

Nested sampling is an alternative to MCMC which gets round some of the issues with standard MCMC techniques, making it useful for parameter estimation and the calculation of the Bayesian evidence. For details on nested sampling, see Skilling (2004).

The development of nested sampling algorithms has taken place within astronomy. Mukherjee et al. (2006) introduced ellipsoidal sampling to improve the sampling procedure while further improvements were made by Feroz & Hobson (2007) and Feroz et al. (2009) who introduced multimodal nested sampling with their routine *Multinest*.

The second paper in this thesis uses *Multinest* to calculate the Bayesian evidence for model selection. Contributing work to the third and fourth paper also use *Multinest* for parameter estimation and calculation of the Bayesian evidence.

### 1.4.3 Radiative transfer

The third and forth paper in this thesis, uses the 1D radiative transfer code, RADEX (van der Tak et al., 2007) combined with the multimodal nested sampling routine *Multinest*, to model spectral line energy distributions (SLEDs) of CO and water. This section

gives a brief overview on radiative transfer and the RADEX code.

The goal of analysing molecular line spectra with radiative transfer codes is to obtain physical parameters of the emitting molecular gas. However, radiative transfer can have numerous levels of complexity. The simplest approach to estimating parameters from SLEDs is to use the 'rotation diagram'. This method assumes the emission is from molecular gas at a single temperature, is excited locally and with energy levels populated according to the Boltzmann distribution, which holds as long as the gas is at a high enough density and low enough optical depth such that collisions dominate over radiation in the excitation process. If this is true, the gas is said to be in local thermodynamic equilibrium (LTE). By assuming LTE, the excitation temperature ($T_{ex}$) and column density of the molecule can be obtained by fitting a straight line to the relative line intensities as a function of upper level energy.

The next level of complexity often referred to as 'non-LTE' radiative transfer, retains the assumption of a local excitation but drops the assumption that the radiation field has very little interaction with the molecular gas and that population levels are thermalised (i.e. energy level populations have a Boltzmann distribution). Since the majority of gas in the interstellar medium is at densities too low for the LTE approximations to be valid, non-LTE models are more appropriate in most circumstances. The radiative transfer equation is solved by assuming the energy levels are in statistical equilibrium, i.e. there is a balance between excitation and de-excitation of molecular energy levels.

Because the molecular level populations and the local radiation field are dependent on each other, the radiative transfer equations forms a recursive problem which has to be solved iteratively. To simplify the calculations, most non-LTE programs use one of two methods. The first is the large velocity gradient (LVG) approach which assumes the cloud has a significant velocity gradient, thereby localising the radiative transfer problem (e.g. Sobolev, 1960; Castor, 1970). The alternative approach is to define a geometrically averaged 'escape probability' ($\beta$) of a photon leaving the gas cloud, based on the current local estimate for the optical depth and an assumed geometry. The level populations calculation now requires a $\beta$ value, which is dependent on geometry and optical depth (Mihalas, 1978).

As the gas is no longer assumed to be in thermodynamic equilibrium, molecular collision data is required. However, if accurate collision rates are known, then column density, kinetic temperature and volume density can all be constrained with the non-LTE approach. The RADEX code (van der Tak et al., 2007) is an example of a non-LTE based

radiative transfer program.

The most sophisticated approach to radiative transfer no longer assumes local excitation and molecular excitation is solved as a function of position in the cloud. These methods make use of numerical (e.g. accelerated Lambda iteration) and Monte Carlo based approaches to iteratively solve radiative transfer calculations for 1D, 2D (e.g. Hogerheijde & van der Tak, 2000) and more recently 3D (e.g. LIME,Brinch & Hogerheijde (2010)) geometries.

The level of complexity required for analysis of molecular line observations is very much dependent on the type of data being analysed. For multiline observations of resolved regions within the Milky Way or nearby galaxies, the 3D, non-local radiative transfer codes such as LIME are the most appropriate. However, extragalactic observations of molecular lines are normally unresolved. This restricts the amount of data to only a handful of lines which are assumed to probe the globally averaged properties of the emitting gas. For these types of problems, the simpler non-LTE based radiative transfer codes are the more appropriate.

Paper three in this thesis contains work using the non-LTE based RADEX code to constrain the physical parameters of molecular gas from observations of the CO SLED in the nearby spiral galaxy IC342. Unlike other approaches that use a grid based $\chi^2$ approach to fit the CO SLEDs (e.g. Rangwala et al., 2011), I use the nested sampling routine, *Multinest*, discussed in Section 1.4.2 to fully sample the posterior parameter space and calculate the Bayesian evidence in order to compare one and two component model fits.

I first fit a one component model to the CO SLED, but find the lowest CO line is underestimated. A two component model provides a better fit, with one component placing tight constraints on the warm gas, and the second component placing looser constraints on a colder gas component.

Paper four includes modelling of $H_2O$ excitation ladder from a starburst galaxy detected at a redshift of 6.34, using a similar setup of RADEX and Multinest. Having carried out various analyses of the $H_2O$ excitation ladder with my modelling setup, it became clear that collisional excitation alone is unable to explain the data and that other mechanisms must play a significant role, for example the pumping of population levels through absorption of photons from the infrared radiation field in star forming regions.

## 1.5 Summary of this Thesis

I have given a brief introduction into the subject of galaxy evolution in the context of the Universe as we know it today. In particular, I have focussed on the processes of star formation and AGN, how they occur, some of the basic signatures of their activity and how they impact on statistical measures such as the luminosity function and cosmic star formation rate density.

I have discussed how observations in the infrared are essential in diagnosing contributions from AGN and star formation. Spectroscopic observations are the most powerful because they contain detailed spectral features such as emission and absorption lines.

The work in this thesis is geared towards applying novel statistical techniques and sophisticated Bayesian inference tools for modelling. I have introduced the basics of these statistical techniques and introduce the framework in which they can be applied to mid-infrared spectroscopy to help untangle the contributions from AGN and star formation. I have also discussed the basics of Bayesian inference and its application to modelling the CO and $H_2O$ SLEDs with radiative transfer models.

The main contents of this thesis is presented in the form of four published papers. Papers one and two show how multivariate techniques provide a more robust analysis of IRS spectra. In particular, they show how the entire mid-infrared region can be used in diagnosing the main power source rather than the present day approach of using only a few spectral features. In paper two, I show that by applying an appropriate technique, these techniques can also be used to blindly learn the spectral components of galaxies. These components are physically intuitive and can be associated with physical processes such as AGN activity and star formation, making them ideal spectral components for decomposing mid-infrared luminosity functions.

In paper three, my Bayesian modelling setup of RADEX and *Multinest* is used to model the CO SLED of IC342. The modelling shows how a one component model is inadequate, whilst a two component model can provide useful constraints on the warm gas component associated with the nuclear activity of the galaxy.

The fourth paper also makes use of the RADEX and *Multinest* setup, in order to test whether the simple model of collisional excitation alone, can explain the $H_2O$ lines observed in one of the most distant starburst galaxies ever detected, HFLS3.

In chapter 6, I summarise all my findings and discuss ongoing and possible future work.

# Bibliography

Barger A. J., Cowie L. L., Sanders D. B., Fulton E., Taniguchi Y., Sato Y., Kawara K., Okuda H. (1998). Submillimetre-wavelength detection of dusty star-forming galaxies at high redshift. *Nature*, 394:248. Cited on 10

Bennett C. L., et al. (2003). First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results. ApJS, 148:1–27, astro-ph/0302207. Cited on 2

Blandford R. D., Payne D. G. (1982). Hydromagnetic flows from accretion discs and the production of radio jets. *Monthly Notices of the Royal Astronomical Society*, 199:883. A&AA ID. AAA031.062.123. Cited on 8

Blanton M. R., Roweis S. (2007). K-corrections and filter transformations in the ultraviolet, optical, and near-infrared. *The Astronomical Journal*, 133:734. Cited on 17

Brandt W. N., Hasinger G. (2005). Deep extragalactic x-ray surveys. *Annual Review of Astronomy & Astrophysics*, 43:827. Cited on 8

Brinch C., Hogerheijde M. R. (2010). Lime - a flexible, non-lte line excitation and radiation transfer method for millimeter and far-infrared wavelengths. *Astronomy and Astrophysics*, 523:25. Cited on 20

Bromley B. C., Press W. H., Lin H., Kirshner R. P. (1998). Spectral classification and luminosity function of galaxies in the las campanas redshift survey. *The Astrophysical Journal*, 505:25. Cited on 14

Burgarella D., et al. (2013). Herschel pep/hermes: the redshift evolution ($0z4$) of dust attenuation and of the total (uv+ir) star formation rate density. *Astronomy & Astrophysics*, 554:70. Cited on 6

Calzetti D., et al. (2005). Star formation in ngc 5194 (m51a): The panchromatic view from galex to spitzer. *The Astrophysical Journal*, 633:871. Cited on 4

Camenzind M. (1990). Magnetized disk-winds and the origin of bipolar outflows. *Reviews in Modern Astronomy*, 3:234. Cited on 8

Castor J. I. (1970). Spectral line formation in wolf-rayet envelopes. *Monthly Notices of the Royal Astronomical Society*, 149:111. A&AA ID. AAA003.114.113. Cited on 19

Condon J. J. (1992). Radio emission from normal galaxies. ARA&A, 30:575–611. Cited on 4

Connolly A. J., Szalay A. S., Bershady M. A., Kinney A. L., Calzetti D. (1995). Spectral classification of galaxies: an orthogonal approach. *Astronomical Journal v.110*, 110:1071. Cited on 14

Dole H., et al. (2006). The cosmic infrared background resolved by spitzer. contributions of mid-infrared galaxies to the far-infrared background. *Astronomy and Astrophysics*, 451:417. Cited on 11

Draine B. T., Li A. (2001). Infrared emission from interstellar dust. i. stochastic heating of small grains. ApJ, 551:807–824. Cited on 6

Eales S., Lilly S., Webb T., Dunne L., Gear W., Clements D., Yun M. (2000). The canada-uk deep submillimeter survey. iv. the survey of the 14 hour field. *The Astronomical Journal*, 120:2244. Cited on 10

Efstathiou A., Rowan-Robinson M. (1995). Dusty discs in active galactic nuclei. MNRAS, 273:649–661. Cited on 10, 14

Fabian A. C. (2012). Observational evidence of active galactic nuclei feedback. *Annual Review of Astronomy and Astrophysics*, 50:455. Cited on 10

Farrah D., et al. (2007). High-resolution mid-infrared spectroscopy of ultraluminous infrared galaxies. *The Astrophysical Journal*, 667:149. Cited on 12

Farrah D., et al. (2008). The nature of star formation in distant ultraluminous infrared galaxies selected in a remarkably narrow redshift range. *The Astrophysical Journal*, 677:957. Cited on 12

Farrah D., et al. (2009). An evolutionary paradigm for dusty active galaxies at low redshift. *The Astrophysical Journal*, 700:395. Cited on 12

Fernandes R. C., Heckman T., Schmitt H., Delgado R. M. G., Storchi-Bergmann T. (2001). Empirical diagnostics of the starburst-agn connection. *The Astrophysical Journal*, 558:81. Cited on 10

Feroz F., Hobson M. P. (2007). Multimodal nested sampling: an efficient and robust alternative to mcmc methods for astronomical data analysis. *arXiv*, astro-ph, 0704.3704v3. Cited on 18

Feroz F., Hobson M. P., Bridges M. (2009). Multinest: an efficient and robust bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398:1601. Cited on 18

Floc'h E. L., et al. (2005). Infrared luminosity functions from the chandra deep field-south: The spitzer view on the history of dusty star formation at $0 <\sim z <\sim 1$. *The Astrophysical Journal*, 632:169. Cited on 10

Fox M. J., et al. (2002). The scuba 8-mjy survey - ii. multiwavelength analysis of bright submillimetre sources. *Monthly Notices of the Royal Astronomical Society*, 331:839. Cited on 10

Genzel R., et al. (1998). What powers ultraluminous iras galaxies? *Astrophysical Journal v.498*, 498:579. Cited on 12

Gulkis S., Lubin P. M., Meyer S. S., Silverberg R. F. (1990). The Cosmic Background Explorer. *Scientific American*, 262:132–139. Cited on 2

Hogerheijde M. R., van der Tak F. F. S. (2000). An accelerated monte carlo method to solve two-dimensional radiative transfer and molecular excitation. with applications to axisymmetric models of star formation. A&A, 362:697–710. Cited on 20

Hopkins P. F., Somerville R. S., Hernquist L., Cox T. J., Robertson B., Li Y. (2006). The relation between quasar and merging galaxy luminosity functions and the merger-driven star formation history of the universe. *The Astrophysical Journal*, 652:864. Cited on 6

Houck J. R., et al. (2004). The infrared spectrograph (irs) on the spitzer space telescope. *The Astrophysical Journal Supplement Series*, 154:18. Cited on 12

Hughes D. H., et al. (1998). High-redshift star formation in the hubble deep field revealed by a submillimetre-wavelength survey. *Nature*, 394:241. Cited on 10

Kauffmann G., et al. (2003). The host galaxies of active galactic nuclei. *Monthly Notices of the Royal Astronomical Society*, 346:1055. Cited on 10

Kennicutt R. C. (1998). Star formation in galaxies along the hubble sequence. *Annual Review of Astronomy and Astrophysics*, 36:189. Cited on 4

Kessler M. F., et al. (1996). The infrared space observatory (iso) mission. *Astron. Astrophys.*, 315:L27. Cited on 12

Khachikian E. Y., Weedman D. W. (1974). An atlas of seyfert galaxies. *Astrophysical Journal*, 192:581. A&AA ID. AAA012.158.104. Cited on 7

Kreckel H., Bruhns H., Čížek M., Glover S. C. O., Miller K. A., Urbain X., Savin D. W. (2010). Experimental results for $h_2$ formation from h and h and implications for first star formation. *Science*, 329(5987):69–71, http://www.sciencemag.org/content/329/5987/69.full.pdf. Cited on 3

Lee D., Seung S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791. Cited on 15, 16

Lee D., Seung S. (2001). Algorithms for non-negative matrix factorization. In *In NIPS*, pages 556–562. MIT Press. Cited on 17

Lemaître G. (1927). Un univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Societe Scientifique de Bruxelles*, 47:49. Cited on 1

Li A., Draine B. T. (2001). Infrared emission from interstellar dust. ii. the diffuse interstellar medium. ApJ, 554:778–802. Cited on 6

Lilly S. J., Fevre O. L., Hammer F., Crampton D. (1996). The canada-france redshift survey: The luminosity density and star formation history of the universe to z approximately 1. *Astrophysical Journal Letters v.460*, 460:L1. Cited on 6

Madau P., Ferguson H. C., Dickinson M. E., Giavalisco M., Steidel C. C., Fruchter A. (1996). High-redshift galaxies in the hubble deep field: colour selection and star formation history to $z \sim 4$. *Monthly Notices of the Royal Astronomical Society*, 283:1388. Cited on 6

Madau P., Pozzetti L., Dickinson M. (1998). The star formation history of field galaxies. *Astrophysical Journal v.498*, 498:106. Cited on 6

Mather J. C. (1982). The Cosmic Background Explorer /COBE/. *Optical Engineering*, 21:769–774. Cited on 2

Matthews T. A., Sandage A. R. (1963). Optical identification of 3c 48, 3c 196, and 3c 286 with stellar objects. *Astrophysical Journal*, 138:30. Cited on 7

Meijerink R., Spaans M. (2005). Diagnostics of irradiated gas in galaxy nuclei. i. a far-ultraviolet and x-ray dominated region code. *Astronomy and Astrophysics*, 436:397. Cited on 5, 10

Meijerink R., Spaans M., Israel F. P. (2007). Diagnostics of irradiated dense gas in galaxy nuclei. ii. a grid of xdr and pdr models. *Astronomy and Astrophysics*, 461:793. Cited on 5

Mihalas D. (1978). Stellar atmospheres /2nd edition/. *San Francisco*. Cited on 19

Mo H., van den Bosch F., White S. (2010). *Galaxy Formation and Evolution*. Galaxy Formation and Evolution. Cambridge University Press. Cited on 7

Mukherjee P., Parkinson D., Liddle A. R. (2006). A nested sampling algorithm for cosmological model selection. *The Astrophysical Journal*, 638:L51. Cited on 18

Murakami H., et al. (2007). The Infrared Astronomical Mission AKARI. PASJ, 59:369, 0708.1796. Cited on 12

Nenkova M., Ivezić Ž., Elitzur M. (2002). Dust emission from active galactic nuclei. *The Astrophysical Journal*, 570:L9. Cited on 10

Neugebauer G., et al. (1984). The infrared astronomical satellite (iras) mission. *Astrophysical Journal*, 278:L1. Cited on 12

Penzias A. A., Wilson R. W. (1965). A measurement of excess antenna temperature at 4080 mc/s. *Astrophysical Journal*, 142:419. Cited on 1

Petric A. O., et al. (2011). Mid-infrared spectral diagnostics of luminous infrared galaxies. *The Astrophysical Journal*, 730:28. Cited on 12

Pilbratt G. L., et al. (2010). Herschel space observatory. an esa facility for far-infrared and submillimetre astronomy. *Astronomy and Astrophysics*, 518:L1. Cited on 12

Planck Collaboration, et al. (2011). Planck early results. I. The Planck mission. A&A, 536:A1, 1101.2022. Cited on 2

Planck Collaboration, et al. (2013). Planck 2013 results. I. Overview of products and scientific results. *ArXiv e-prints*, 1303.5062. Cited on 2

Rangwala N., et al. (2011). Observations of arp 220 using herschel-spire: An unprecedented view of the molecular gas in an extreme star formation environment. *The Astrophysical Journal*, 743:94. Cited on 20

Rigopoulou D., Spoon H. W. W., Genzel R., Lutz D., Moorwood A. F. M., Tran Q. D. (1999). A large mid-infrared spectroscopic and near-infrared imaging survey of ultraluminous infrared galaxies: Their nature and evolution. *The Astronomical Journal*, 118:2625. Cited on 12

Rowan-Robinson M. (1977). On the unity of activity in galaxies. ApJ, 213:635–647. Cited on 8

Rowan-Robinson M., et al. (1997). Observations of the hubble deep field with the infrared space observatory - v. spectral energy distributions, starburst models and star formation history. *Monthly Notices of the Royal Astronomical Society*, 289:490. Cited on 10

Salim S., et al. (2007). Uv star formation rates in the local universe. *The Astrophysical Journal Supplement Series*, 173:267. Cited on 4

Salpeter E. E. (1964). Accretion of interstellar matter by massive objects. *Astrophysical Journal*, 140:796. Cited on 8

Sanders D. B. (1999). A new view of galaxy evolution from submillimeter surveys with scuba. *Astrophysics and Space Science*, 269:381. Cited on 10

Seyfert C. K. (1943). Nuclear emission in spiral nebulae. *Astrophysical Journal*, 97:28. Cited on 7

Siebenmorgen R., Krügel E. (2007). Dust in starburst nuclei and ulirgs. sed models for observers. *Astronomy and Astrophysics*, 461:445. Cited on 14

Skilling J. (2004). Nested sampling. *BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. AIP Conference Proceedings*, 735:395. Cited on 18

Sobolev V. V. (1960). Moving envelopes of stars. *Cambridge: Harvard University Press.* Cited on 19

Soifer B. T., et al. (1984). Infrared galaxies in the iras minisurvey. *Astrophysical Journal*, 278:L71. Cited on 12

Spoon H. W. W., Marshall J. A., Houck J. R., Elitzur M., Hao L., Armus L., Brandl B. R., Charmandaris V. (2007). Mid-infrared galaxy classification based on silicate obscuration and pah equivalent width. *The Astrophysical Journal*, 654:L49. Cited on 12

Taghizadeh-Popp M., Heinis S., Szalay A. S. (2012). Single parameter galaxy classification: The principal curve through the multi-dimensional space of galaxy properties. *The Astrophysical Journal*, 755:143. Cited on 14

Tielens A. G. G. M. (2005). *The Physics and Chemistry of the Interstellar Medium*. Cited on 5

Urry C. M., Padovani P. (1995). Unified schemes for radio-loud active galactic nuclei. *Publications of the Astronomical Society of the Pacific*, 107:803. Cited on 8, 9

Valiante E., Lutz D., Sturm E., Genzel R., Chapin E. L. (2009). A backward evolution model for infrared surveys: The role of agn- and color-l tir distributions. *The Astrophysical Journal*, 701:1814. Cited on 11

van der Tak F. F. S., Black J. H., Schöier F. L., Jansen D. J., van Dishoeck E. F. (2007). A computer program for fast non-lte analysis of interstellar line spectra. with diagnostic plots to interpret observed line intensity ratios. *Astronomy and Astrophysics*, 468:627. Cited on 18, 19

Wang L., Farrah D., Connolly B., Connolly N., LeBouteiller V., Oliver S., Spoon H. (2011). Principal component analysis of the spitzer irs spectra of ultraluminous infrared galaxies. *Monthly Notices of the Royal Astronomical Society*, 411:1809. Cited on 14

Werner M. W., et al. (2004). The spitzer space telescope mission. *The Astrophysical Journal Supplement Series*, 154:1. Cited on 12

Wright E. L., et al. (2010). The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. AJ, 140:1868–1881, 1008.0031. Cited on 12

Wu Y., et al. (2011). The mid-infrared luminosity function at z ¡ 0.3 from 5muses: Understanding the star formation/active galactic nucleus balance from a spectroscopic view. *The Astrophysical Journal*, 734:40. Cited on 11

# Chapter 2

# Principal Component Analysis and Radiative Transfer modelling of Spitzer IRS Spectra of Ultra Luminous Infrared Galaxies

*Peter Hurley, Seb Oliver, Duncan Farrah, Lingyu Wang, Andreas Efstathiou*

## 2.1 Abstract

The mid-infrared spectra of ultraluminous infrared galaxies (ULIRGs) contain a variety of spectral features that can be used as diagnostics to characterise the spectra. However, such diagnostics are biased by our prior prejudices on the origin of the features. Moreover, by using only part of the spectrum they do not utilise the full information content of the spectra. Blind statistical techniques such as principal component analysis (PCA) consider the whole spectrum, find correlated features and separate them out into distinct components.

We further investigate the principal components (PCs) of ULIRGs derived in Wang et al. (2011). We quantitatively show that five PCs is optimal for describing the IRS spectra. These five components (PC1-PC5) and the mean spectrum provide a template basis set that reproduces spectra of all $z < 0.35$ ULIRGs within the noise. For comparison, the spectra are also modelled with a combination of radiative transfer models of both starbursts and the dusty torus surrounding active galactic

nuclei. The five PCs typically provide better fits than the models. We argue that the radiative transfer models require a colder dust component and have difficulty in modelling strong PAH features.

Aided by the models we also interpret the physical processes that the principal components represent. The third principal component is shown to indicate the nature of the dominant power source, while PC1 is related to the inclination of the AGN torus.

Finally, we use the 5 PCs to define a new classification scheme using 5D Gaussian mixtures modelling and trained on widely used optical classifications. The five PCs, average spectra for the four classifications and the code to classify objects are made available at: `http://www.phys.susx.ac.uk/~pdh21/PCA/`.

## 2.2 Introduction

Ultraluminous Infrared Galaxies (ULIRGs) are galaxies whose rest-frame infrared luminosities, $L_{8-1000\mu\mathrm{m}}$, exceed $10^{12}\mathrm{L}_{\odot}$. Although ULIRGs were first discovered using ground based photometry in the 1970s (Rieke & Low, 1972), the *IRAS* survey transformed our understanding by observing the objects in much larger numbers (Soifer et al., 1984). Most have high star-formation rates ($SFR > 100\mathrm{M}_{\odot}\mathrm{yr}^{-1}$), while around half also contain an embedded Active Galactic Nucleus (AGN).

ULIRGs are rare in the local Universe, with less than fifty at $z \lesssim 0.1$, but the associated luminosity function shows strong, positive evolution with redshift (e.g. Sanders, 1999), resulting in several hundred ULIRGs per square degree at $z > 1$ (Rowan-Robinson et al., 1997; Barger et al., 1998; Hughes et al., 1998; Eales et al., 2000; Fox et al., 2002; Le Floc'h et al., 2005). The increase in number density with redshift and their associated high SFR means ULIRGs make a significant contribution to the history of star formation at high $z$.

The mid to far infrared luminosity of ULIRGs is a result of dust and gas reprocessing the optical and UV radiation emitted by stars and/or AGN. Obtaining spectroscopy for the mid-infrared part of the spectrum became possible with instruments such as the *Infrared Space Observatory* (*ISO*; Kessler et al. (1996)), and the Infrared Spectrograph (*IRS*; Houck et al. (2004)) on the *Spitzer Space Telescope* (Werner et al., 2004). The ULIRG spectra from these instruments contain a wealth of spectral features. These include the emission lines from broad polycyclic aromatic

hydrocarbons (PAHs), which are strong in starforming regions, but absent in AGN dominated sources (Moorwood, 1986; Roche et al., 1991). A prominent [Ne V] 14.3 $\mu$m fine structure line indicates the presence of an AGN, while the silicate features at 9.7 and 18 $\mu$m probe source geometry (Imanishi et al., 2007).

Combinations of the PAH emission lines, mid-infrared fine-structure lines and silicate features have been used as diagnostics for characterising the power source behind the ULIRGs (Genzel et al., 1998; Rigopoulou et al., 1999; Spoon et al., 2007; Farrah et al., 2007, 2008, 2009). There are however problems associated with these diagnostic tools, such as the separation of emission lines from both the continuum and underlying PAH features, the mixture of neighbouring features and different diagnostics giving conflicting estimates. They also only focus on small parts of the spectrum, disregarding the information contained in the remainder.

Larger regions of the spectrum can be investigated with the multivariate statistic, Principal Component Analysis (PCA). PCA has been used for spectral classification for optical galaxies (e.g. Connolly et al., 1995; Bromley et al., 1998). Wang et al. (2011) carried out PCA on the *IRS* spectra of 119 local ULIRGs. They argued, qualitatively, that only 4 principal components (PCs) were needed to reproduce the variance in the ULIRG spectra. They also proposed that the contribution from each PC had some underlying physical interpretation. Examination of the first four PCs, and comparisons to the diagnostics employed by Spoon et al. (2007) and Nardini et al. (2009) suggested that PC1 constrains the dust temperature and geometry of the distribution of source and dust, while PC2 and PC3 determine the amount of star formation. The fourth PC is important for Seyfert Type 2 galaxies, and is hence a possible indicator of an unobscured AGN.

In this paper we extend Wang et al. (2011) by quantitatively investigating how many PCs are needed to explain the variation in the spectra and compare the PC reconstructions to fits provided by a suite of radiative transfer models. We investigate what information the radiative transfer models are missing. We also re-examine what physical properties are behind the PCs, by investigating the relationship between the physical parameters of models and the contributions from different PCs. Finally, we introduce a new classification scheme using 5D Gaussian mixtures modelling and trained with optical classifications. Section 2.3 gives an overview of the data and Section 2.4 a brief description of PCA. Section 2.5 will

review the radiative transfer models being applied, and Section 2.6 will present the results. Conclusions will be presented in Section 2.7. We assume a spatially flat cosmology with $H_0 = 70 \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}, \Omega = 1$, and $\Omega_m = 0.3$.

## 2.3 The Data

This paper uses the same sample of mid-infrared spectra as Wang et al. (2011). We summarise their selection criteria here. The ULIRGs were observed as part of the IRS Guaranteed Time program (Armus et al., 2007; Farrah et al., 2007; Spoon et al., 2007) and those observed by Imanishi et al. (2007). An upper redshift cut of z = 0.35 was applied to ensure we sample approximately the same wavelength range for each object. A further eight objects were removed as they have poor-quality data in the longer-wavelength IRS module. In total, there are 119 objects in the sample.

## 2.4 Principal Component Analysis (PCA)

PCA works by determining the eigenvectors from the covariance matrix of a given dataset. For 119 spectra, each with 180 wavelength points, the 180 by 180 covariance matrix quantifies the correlation between each spectral point. The eigenvectors of the matrix can be thought of as spectral components that can be linearly combined to reconstruct each object in the sample.

Any spectrum can be linearly decomposed by projecting it onto the principal components defined by the 119 ULIRG sample. This allows each spectrum to be described by the contribution from each PC. These contributions define co-ordinates in a multidimensional space which we refer to as PCA space.

## 2.5 Radiative transfer models

To compare with the fits provided by the principal components, we have carried out a minimum chi squared search for linear combinations of a grid of starburst models described in Siebenmorgen & Krügel (2007) and grid of AGN dusty torus models of Efstathiou & Rowan-Robinson (1995). The libraries contain 5948 and 2109 SEDS respectively and we have considered linear combinations of each AGN and starbust SED, giving us many models to search over.

| Parameter | Range |
|---|---|
| $R$ (kpc) | 0.35, 1 and 3 |
| $f_{OB}$ | 0.4, 0.6 and 0.9 |
| $L_{SB}$ ($L_\odot$) | $10^{10}$ to $10^{14}$ in steps of 0.1 dex |
| $A_v$ (Mag) | 2.2, 4.5, 7, 9, 18, 35, 70 and 120 |
| $n_{HS}$ ($cm^{-3}$) | $10^2, 10^3, 2.5 \cdot 10^3, 5 \cdot 10^3, 7.5 \cdot 10^3, 10^4$ |

Table 2.1: Parameter values and ranges for the starburst models.

### 2.5.1 Starburst Models

We use the Siebenmorgen & Krügel (2007) starburst models. The models presented by Siebenmorgen & Krügel (2007) have been described as 'hot spot' starbursts. OB stars are assumed to be surrounded by dense clouds (the hot spots) and other stars, such as old bulge stars or massive stars are dispersed in the diffuse medium. It is the hot spots that contribute to the mid infrared part of the spectrum. The outer radius of these environments is determined by the condition of equal heating of the dust by the OB stars in the centre and the interstellar radiation field.

Both stellar groups are treated as continuously distributed sources, and the number density of both types of stars, falls off as $r^{-1}$.

The parameters of these models include the starburst radius, $R$; ratio of the luminosity of OB stars with hot spots to total luminosity, $f_{OB}$; the total luminosity of the starburst, $L_{SB}$; total extinction from the outer radius of the galactic nucleus to its centre, $A_v$; and dust density of the hot spot environment, $\rho_{HS}$, corresponding to hydrogen number densities ($n_{HS}$) and assuming a gas to dust ratio of 150. The parameter ranges can be found in Table 2.1. In total, the library contains 5948 SEDs.

### 2.5.2 AGN torus models

This paper uses the AGN tapered disc models of Efstathiou & Rowan-Robinson (1995). The tapered disc models, in combination with the starburst models of Efstathiou et al. (2000), have been successful in fitting the spectral energy distributions of ultraluminous infrared galaxies (Farrah et al., 2003), hyperluminous infrared galaxies (Farrah et al., 2002; Verma et al., 2002; Efstathiou, 2006) submillimeter galaxies (Efstathiou & Siebenmorgen, 2009), and active galaxies (Alexander

| Parameter | Range |
|---|---|
| $\tau$ | 500, 750, 1000, 1250 |
| $\Theta$ (degrees) | 30, 45, 60 |
| $(r_{\rm in}/r_{\rm out})^{-1}$ | 20, 60, 100 |
| $\theta$ (degrees) | 0 to 90 with either |
| | 40 or 75 divisions (depending on $r_{\rm in}/r_{\rm out}$) |

Table 2.2: Parameter values and ranges for the AGN models.

et al., 1999; Efstathiou & Siebenmorgen, 2005; Farrah et al., 2012; Ruiz et al., 2001). The torus is modelled as a disc, whose thickness increases with distance from the central source but tapers off in the outer regions of the torus. The dust density is distributed smoothly within the disc and follows a $r^{-1}$ relation, with r being radius. The parameters for the AGN torus model are: ultraviolet equatorial optical depth to the centre of the torus, $\tau$; the opening angle of the torus, $\Theta$; the ratio of inner to outer radius of the torus, $(r_{\rm in}/r_{\rm out})^{-1}$; and the viewing angle, $\theta$ (see Table 2.2). In total, there are 2109 AGN SEDs.

### 2.5.3 The fitting procedure

We have considered all linear combinations of a starburst and AGN model when fitting the observed spectra of the 119 ULIRG sample. We use the wavelength grid of the starburst models, and the lower resolution AGN models are interpolated onto the same grid. The smoothness of the AGN models, makes the interpolation justifiable. The radiative transfer models lack molecular hydrogen emission so we mask out regions of the spectrum where molecular hydrogen features occur ( i.e. $9.46 - 9.86, 12.08 - 12.48$ and $16.83 - 17.23\mu$m).

The wavelength resolution of the PCs is higher than the starburst model resolution. For proper comparison to the fits, and to allow decomposition of the models into PCA space, we have re-derived the principal components for the ULIRG sample at the resolution of the starburst models. There is no significant change in the shape of components. We also note that the sign of the PC contributions for each object, remains the same and the change in magnitude of the PC contributions is not significant in comparison to the spread of contributions for the sample.

To remain consistent with the analysis of Wang et al. (2011), the models are

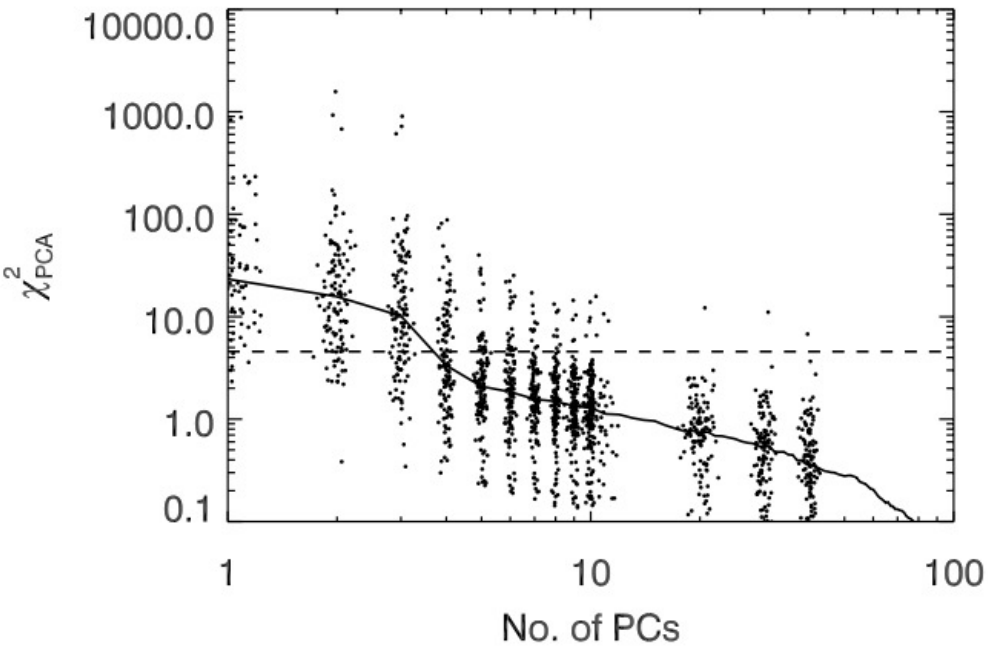


Figure 2.1: The median variation of the $\chi^2_\nu$ for the PC reconstruction as the number of components used in the reconstruction are increased. The dashed line indicates the median $\chi^2_\nu$ for the radiative transfer model fits. For PC reconstructions using up to 10 PCs (and 20, 30, 40) we also plot the $\chi^2_\nu$ for every object (offset for clarity)

normalised so that the mean flux over the whole wavelength range is unity.

We then carry out a linear least squares fit for each combination of starburst and AGN model, with the condition that the fit parameters are positive (i.e. to eliminate the possibility of a negative amount of starburst or AGN). Model comparison is then carried out via minimum chi squared ($\chi^2$).

We assumed a minimum of 5% flux error for each spectral bin of the IRS spectra, which is consistent with the observed variations between individual nod positions on the IRS as described in Chapter 7 of the IRS Instrument Handbook*.

## 2.6 Results

### 2.6.1 Optimum number of components

We first investigate how many PCs are needed to describe the ULIRG sample. Wang et al. (2011) did not quantitatively show whether 4 PCs were sufficient. Using the

*http://irsa.ipac.caltech.edu/data/SPITZER/docs/irs/

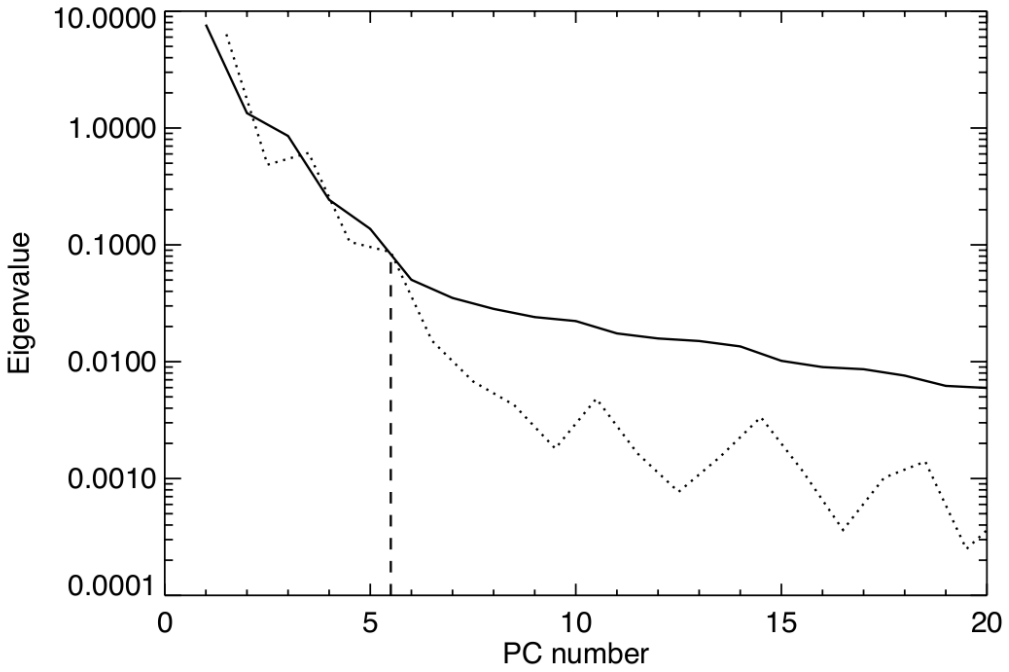


Figure 2.2: The eigenvalues (solid line) and difference in eigenvalues (dotted line) for the PCs. The eigenvalues quantify the variance associated with each PC, and are a measure of importance. The difference between eigenvalues drops dramatically for the first few PCs, but levels off beyond 5 (indicated by the dashed line). We therefore argue that 5 PCs is a more suitable number than the 4 used in Wang et al. (2011).

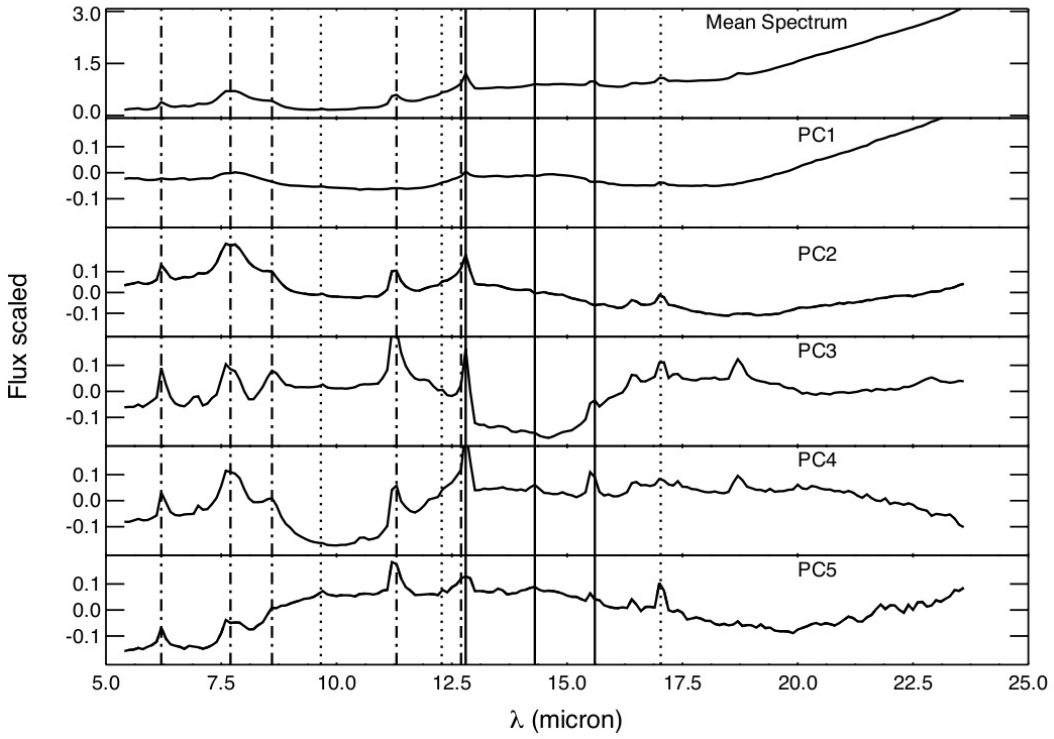


Figure 2.3: The mean spectrum and principal components for the sample of ULIRGs. The dot-dashed vertical lines mark the central location of the 6.2,7.7,8.6,11.2 and 12.7 $\mu$m PAH emission lines. The dotted lines indicate the location of the molecular hydrogen lines at 9.66, 12.28 and 17.03 $\mu$m. The solid vertical lines indicate the position of the neon fine-structure lines, [Ne II] 12.8, [Ne v] 14.3 and [Ne III] 15.6 $\mu$m.

PCs re-derived at the lower resolution described in Section 2.5.3, we have investigated how many PCs are needed to accurately reconstruct the IRS spectra of all 119 ULIRGs in the sample. For each spectrum, we quantify the goodness of reconstruction with the reduced chi squared statistic $\chi^2_\nu$, where the number of degrees of freedom is equal to the number of wavelength points minus the number of PCs used in the reconstruction.

Figure 2.1 shows that as we increase the number of PCs used in the reconstruction, the median $\chi^2_\nu$ value for the sample decreases. We have plotted the $\chi^2_\nu$ for each individual object for reconstructions using up to ten PCs and the median $\chi^2_\nu$ value obtained by fitting the ULIRGs with the radiative transfer models as described in Section 2.5.3. Ten PCs would appear to be the optimal number i.e. where $\chi^2_\nu = 1$. We find that four PCs (assumed by Wang et al. (2011)) give a median $\chi^2_\nu$ of 3.3, while adding a fifth component substantially decreases the median $\chi^2_\nu$ to 2.1. The

use of six and seven PCs only reduces the median $\chi^2_\nu$ to 1.8 and 1.6 respectively.

The eigenvalues associated with each PC are a measure of the variance each PC accounts for and provide an alternative method to determine the optimum number of components. In Figure 2.2, we plot the eigenvalues and difference in eigenvalues for the PCs. The general trend indicates the difference between eigenvalues significantly decreases with each component. The exception to the rule occurs between the 3rd-4th component and the 5th-6th component where the difference between eigenvalues is larger than the trend. We associate this larger than expected difference as in indication that the previous component captures significantly more information than the next. This suggests that the third and fifth PC are substantially more important than the fourth and sixth respectively. Beyond the sixth PC, the trend flattens out, indicating most of the variation related to structure has been captured. Overall, Figures 2.1 and 2.2 do not definitively indicate the optimum number of PCs. However, we argue that the reduction in $\chi^2_\nu$ to 2.1 and difference in eigenvalue between the fifth and sixth PC, indicates that five PCs rather than the four PCs used by Wang et al. (2011), strike a better balance of providing a small basis set of templates, whilst adequately describing the spectra.

The fifth component was not discussed in Wang et al. (2011) and so we now show this component, compared to the original four. The mean spectra of the 119 ULIRGs and the 5 components can be seen in Figure 2.3. There are a number of spectral features in this fifth component, most notably the 6.2, 11.2 and 12.7 $\mu$m PAH emission lines as well as the molecular hydrogen emission line at 17.03 $\mu$m. The 6.2 $\mu$m emission feature has negative flux, while the 11.2 and 12.7 PAH lines are both positive. Overall, the fifth component does not contain any new features that were not seen in the previous components. Its role appears to be in altering the ratios of existing features.

### 2.6.2 Analysis of the radiative transfer models

We now investigate whether the radiative transfer models discussed in Section 2.5 are capable of modelling the spectra. An example of the fit produced by 5 PCs and the radiative transfer models can be seen in Figure 2.4.

We now compare all the $\chi^2_\nu$ for reconstructions using 5 PCs with the $\chi^2_\nu$ for our radiative transfer model fits. Figure 2.5 shows the distribution of the reduced chi

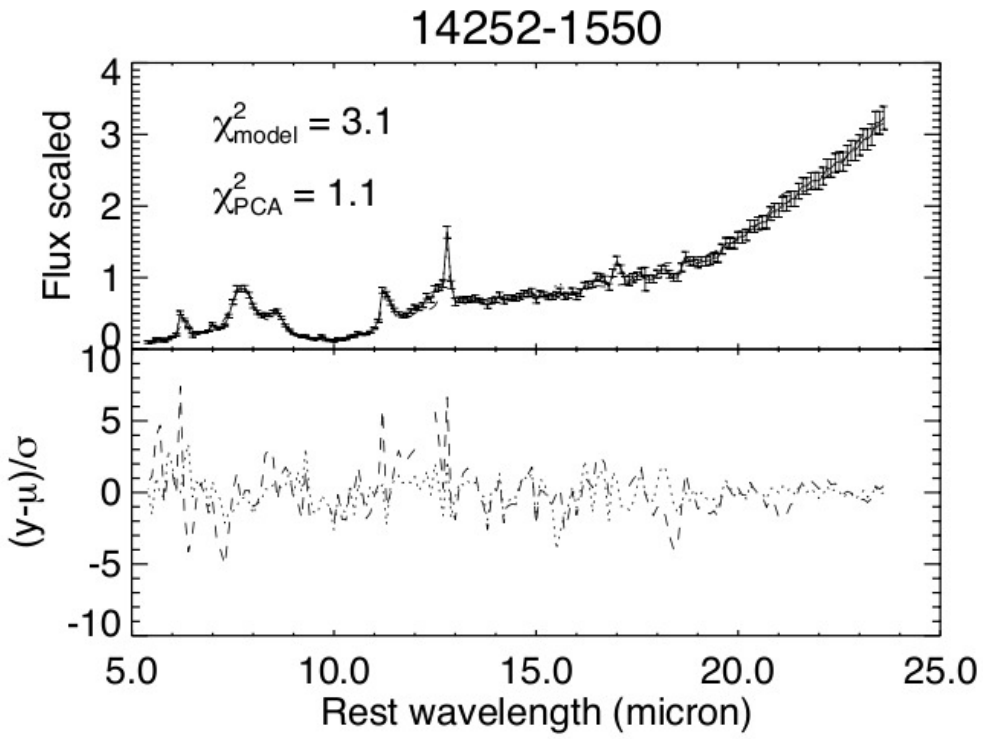


Figure 2.4: An example of our fit with 14252-1550. The radiative transfer model is plotted with a dashed line, and the principal component reconstruction with 5 PCs is shown with a dotted line. The residual over error is also shown to indicate where either technique may be failing

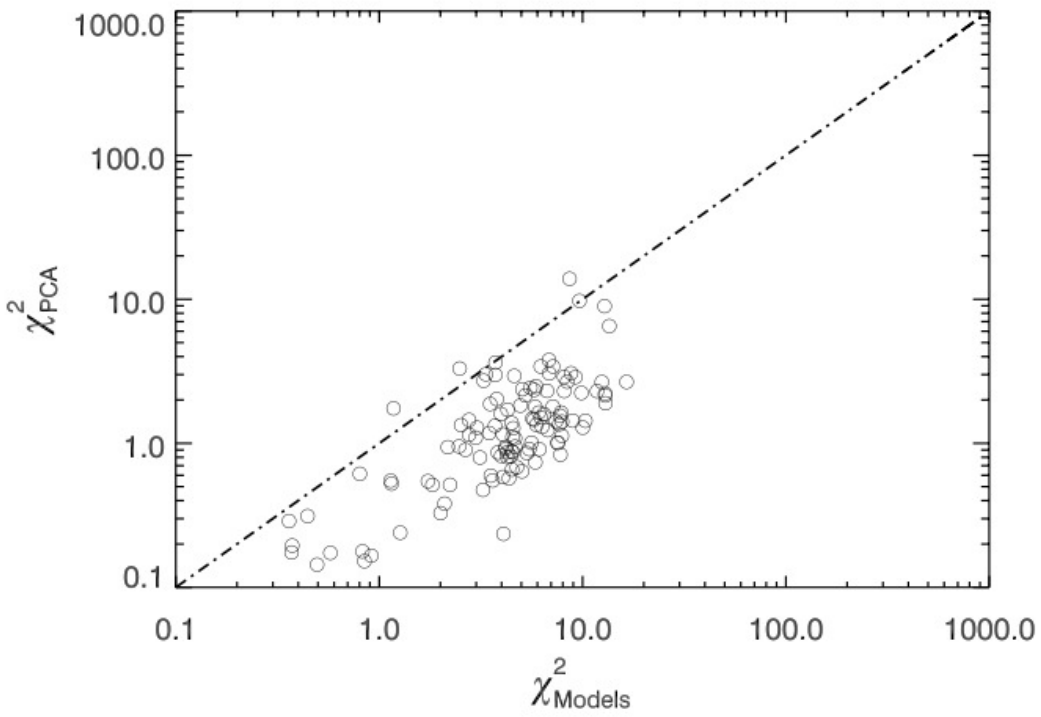


Figure 2.5: The $\chi^2_\nu$ values for each object in the sample for both radiative transfer model fits and the 5 PC reconstruction. Most objects do better with the PCs.

squared values for both the 5 PC reconstructions and the radiative transfer model fits, for all ULIRGs in the sample. A 5 component reconstruction fits the spectra better, on average, than the radiative transfer models.

We have shown that 5 PCs can explain the sample of ULIRGs better than the radiative transfer models, but the two are not competing methodologies. The PCs will always do better than the models as they are derived from the data and the number of PCs is increased until the reproduction of the spectra is good. They represent an extraction of most of the important information from the spectra. Radiative transfer models are used to give us physical information of objects. However, Figure 2.5 indicates that the ULIRGs are not modelled well on average by the radiative transfer models.

By comparing the models to the PCs, we can investigate what information is in the PCs that is not in the models. Figure 2.6 shows the contributions made by the five PCs, as a function of $\chi^2_\nu$ for the model fits. We only plot objects that have a reasonable $\chi^2_\nu$ for the 5 PC reconstruction i.e. a $\chi^2_\nu \leq 3$. We also bin the model $\chi^2_\nu$ values into three bins. The mean and one sigma dispersion are overplotted as filled circles and errorbars.

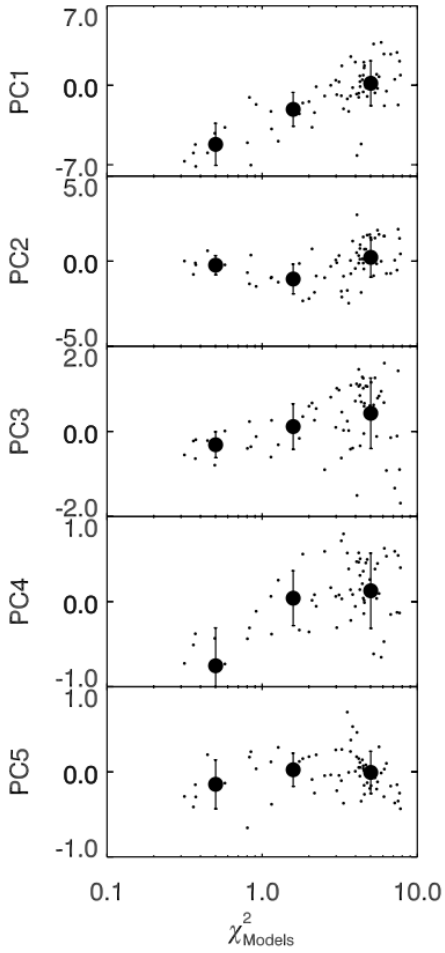


Figure 2.6: The contributions made by each PC against the $\chi^2_\nu$ from the model fits. Only objects with a 5 PC fit of $\chi^2_\nu \leq 3$ have been plotted. The mean and one sigma dispersion for three bins are overplotted as filled circles and errorbars

The general increase of $\chi^2_{\nu Models}$ in Figure 2.6 shows that models tend to do worse when the objects have a large, positive contribution from PC1. Wang et al. (2011) suggested a large, positive contribution from PC1 indicated colder dust. Our results suggest that objects with colder dust are not well modelled by the AGN and starburst component models. The increase in dispersion with $\chi^2_{\nu Models}$ for PC2 and PC3 indicates models do worse when there is a large, absolute contribution from PC2 and PC3. PC2 and PC3 relate to strong spectral lines, which would indicate that the models have problems with constraining the strength of spectral lines. Lower values of $\chi^2_{\nu Models}$ appear to occur when objects have negative values of PC4, but as the $\chi^2_{\nu Models}$ values increase beyond 2, there appears to be little change in PC4. A negative contribution in PC4 would suppress emission features, indicating that models are again inadequate in modelling spectral features. There appears to be little change of PC5 contribution with $\chi^2_{\nu Models}$.

In Figure 2.7, we show the stacked difference between spectra and radiative transfer model fits (solid line) and the spectra and 5 PC reconstructions (dotted line). The stacked difference for spectra and models illustrates that model fits underestimate the PAH spectral lines and do not include Neon fine structure lines, or molecular Hydrogen lines. The PAH underestimate is consistent with our interpretation of Figure 2.6. It suggests the Kruegel (2003) PAH treatment used by the Siebenmorgen & Krügel (2007) starburst models, is unsuitable for the extreme starforming ULIRGs. As expected, the PC reconstructions perform considerably better than the models.

### 2.6.3 Interpreting the Principal Components

We have shown that 5 PCs provide a simple empirical basis set that capture most of the important variations in ULIRGs. We have also shown some limitations of the models. Nevertheless, the models still describe some of the physics of the objects and can be cautiously used to investigate whether the components are associated with physical parameters. We investigate the components by directly comparing the PC contributions and the radiative transfer model best fits for the ULIRG sample. Figure 2.8 shows the contribution from each PC as a function of the viewing angle and starburst/AGN contribution. We have binned the PC contributions and calculated the average and one sigma dispersion for each bin. These are over-plotted with errorbars. PC1 shows a correlation with viewing angle of AGN, with positive

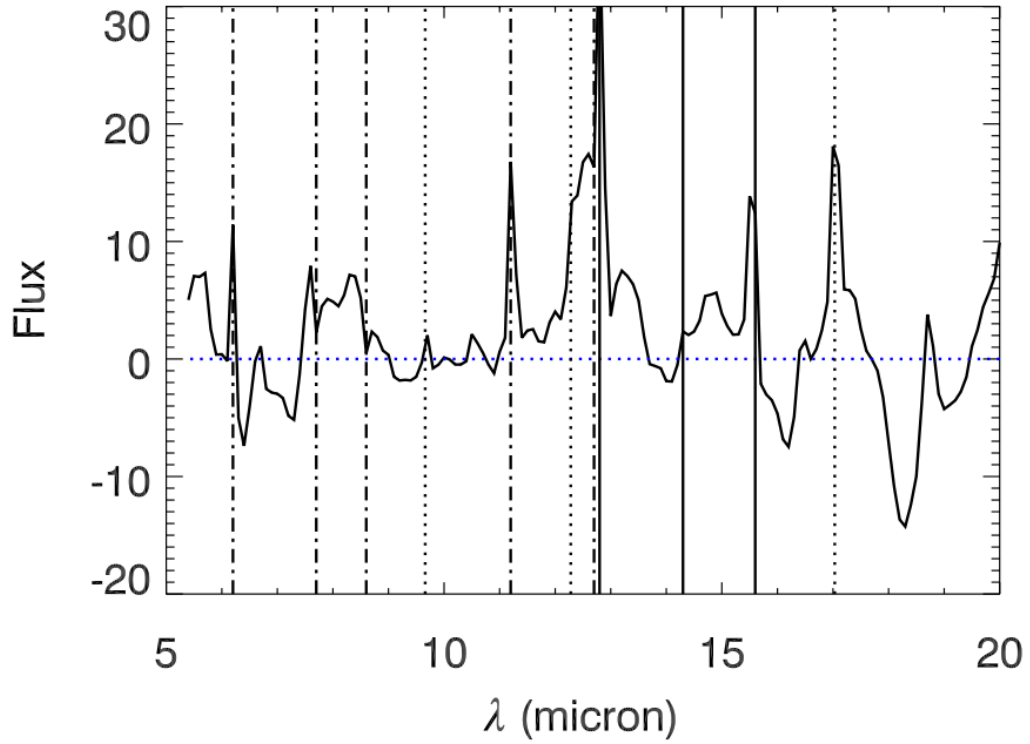


Figure 2.7: The stacked difference between ULIRG spectra and best fit radiative transfer models (solid line) and the ULIRG spectra and 5 PC reconstructions (dotted line). The dot-dashed vertical lines mark the central location of the 6.2,7.7,8.6,11.2 and 12.7 $\mu$m PAH emission lines. The dotted lines indicate the location of the molecular hydrogen lines at 9.66, 12.28 and 17.03 $\mu$m. The solid vertical lines indicate the position of the neon fine-structure lines, [Ne II] 12.8, [Ne v] 14.3 and [Ne III] 15.6 $\mu$m.

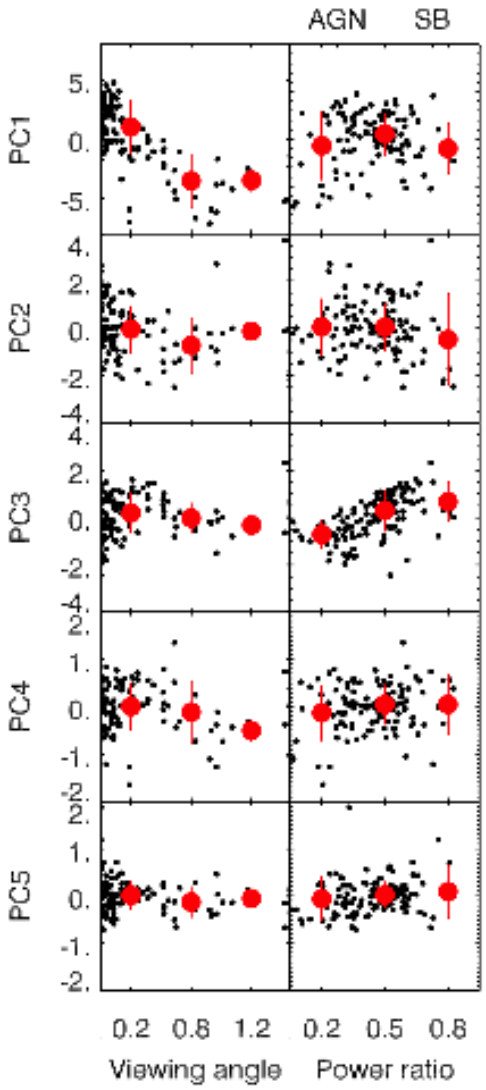


Figure 2.8: The contribution from each PC against the radiative transfer parameters of viewing angle (in radians) and starburst/AGN contribution. The average contribution for three bins and associated one sigma dispersion are overplotted.

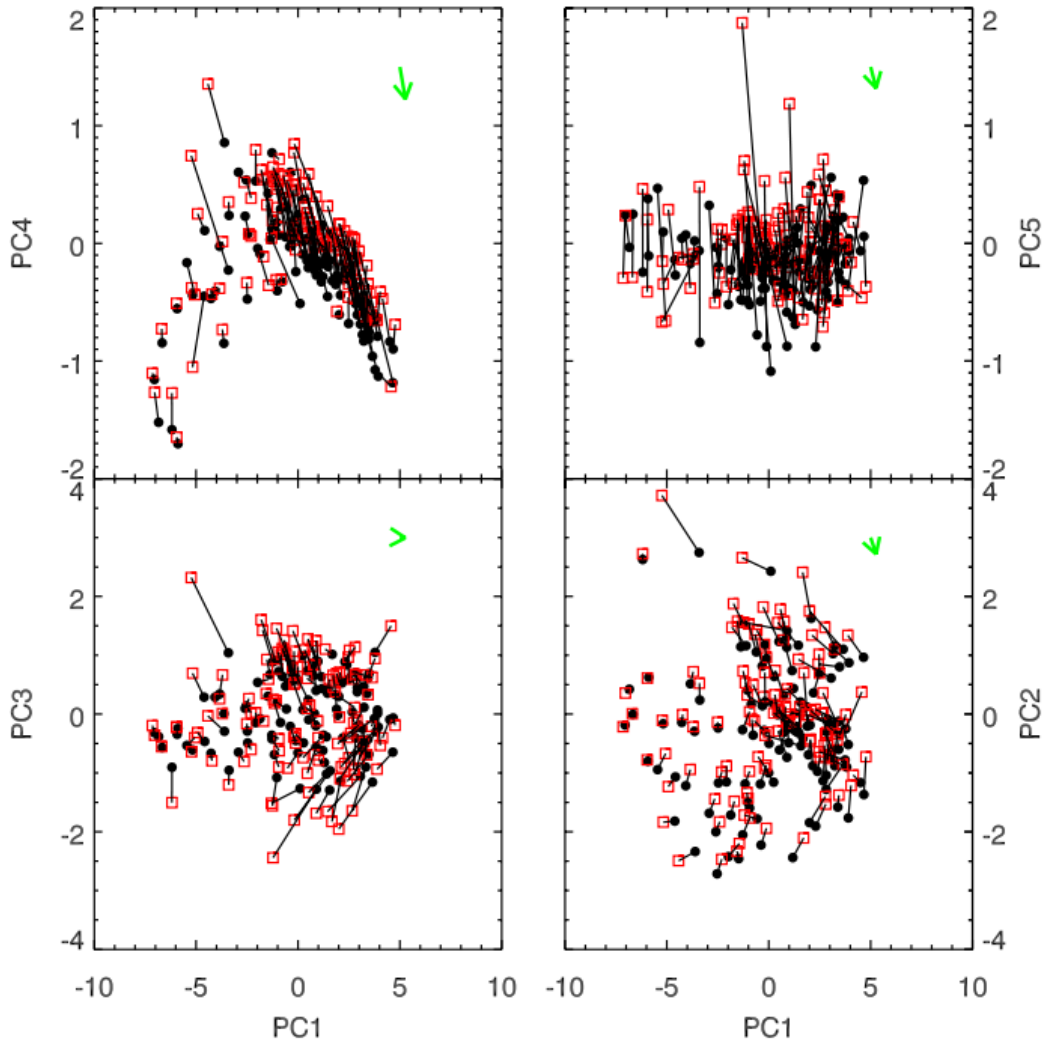


Figure 2.9: The position of ULIRGs in four of the PC planes (squares) and the position in PCA space of the corresponding best fit radiative transfer models (filled circles). Each ULIRG and best fit model are joined by a solid line. The arrows in the top right of each plot show the mean difference between ULIRGs and models.

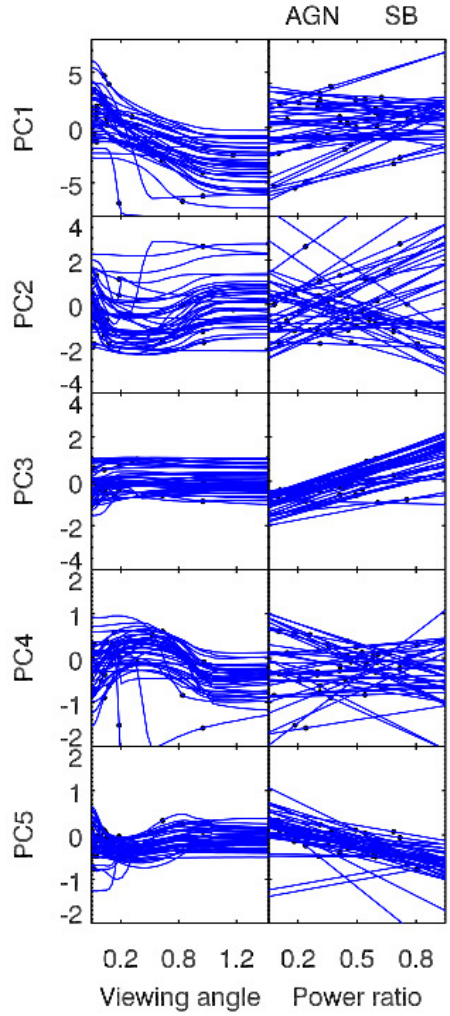


Figure 2.10: The contribution from each PC against viewing angle tracks and power ratio for 50 of the ULIRGs. For each best fit radiative transfer model, the viewing angle and power ratio have been varied to create tracks in PCA space.

contributions corresponding to an obscured AGN and negative to face on AGN. The contribution from the fourth PC appears to drop with viewing angle from around $\frac{\pi}{4}$ radians. The other PCs show no discernible dependence. The starburst/AGN contribution is plotted against PCs in the right hand side of Figure 2.8. Negative values of PC3 seem to be associated with AGN dominated sources and positive values with starbursts. The other PCs show a large amount of dispersion and little correlation with starburst/AGN contribution.

We now decompose the radiative transfer model fits into the PCA space described in Section 2.4. The position of each ULIRG (squares) in four of the PCA planes and corresponding best fit model (filled circles) can be seen in Figure 2.9. The mean difference between the ULIRGs and models is depicted by the arrow in the top right of each plane.

We find the location of radiative transfer model best fits in PCA space are offset relative to the ULIRG positions. There are numerous explanations for the offset. The sparseness of the model library could be a factor. The decomposition into PCA space may also be affected by the missing physics in the models. We therefore treat the model tracks with caution and limit interpretation to relative changes in PC contribution rather than absolute position.

We have taken the best fit radiative transfer model and vary each parameter in turn to see how it affects the position in PCA space. We focus on the viewing angle of AGN and ratio of starburst to AGN power, which we define as:

$$L_{\text{total}} = pL_{\text{SB}} + (1-p)L_{\text{AGN}} \tag{2.1}$$

A value $p = 0$ describes a pure AGN model, and $p = 1$ relates to a complete starburst.

Figure 2.10 shows the 1D parameter tracks for 50 randomly selected ULIRGs. The viewing angle tracks show a decrease in PC1 contribution when going from an obscured to face on AGN. Tracks in PC4 are curved, indicating a non-linear relationship with viewing angle. PC3 appears to be a good indicator for the power ratio, with PC3 contribution decreasing as AGN power begins to dominate. Tracks in PC5 also show a slight correlation with power ratio, while for the other PCs the relationship is unclear.

The interpretation of tracks is consistent with the conclusions drawn from Figure

2.8. Certain PCs appear to be related to the physics of the ULIRGs. We have shown that PC1 is linked to AGN viewing angle, while PC3 is linked to the star formation and AGN contribution. This is consistent with the interpretation of Wang et al. (2011).

### 2.6.4 Gaussian mixtures classification scheme

Since we have shown the PCs capture most of the information in IRS spectra, it is natural to use the PCs as a classification tool. Wang et al. (2011) suggested that position in the PC1-PC4 plane was related to optical type. We now take take this one step further by proposing a classification scheme based on optical classifications, using the multi-dimensional Gaussian mixtures modelling (GMM) applied in Davoodi et al. (2006). This type of parametric modelling works by assuming the density function of galaxies in our 5D PCA space is composed of a mixture of multidimensional Gaussian functions. We take the four optical classifications (Seyfert 1, Seyfert 2, LINER and HII) that exist for 78 of our 119 ULIRG sample, and assume the density of objects in each classification can be described as Gaussian. The resulting position and width of each Gaussian are trained from the optical classifications. They can be thought of as a probability density function (PDF) that describes the probability of belonging to each optical classification, as a function of position in PCA space. See Appendix A for more details on GMM.

Figure 2.11 shows the marginalised one sigma contours for the optical classifications in four 2D projections. We note that the one sigma contours are for visualisation only, our classification scheme makes use of all 5 dimensions. The objects with optical classifications are represented with different symbols: crosses for Seyfert 1, triangles for Seyfert 2, squares for LINER and open circles for objects classified as HII. Objects without an optical classification are plotted with a diamond. The success rate of our classification, can be found in Table 2.3.

The classification scheme is very successful in correctly identifying Seyfert 1 like objects, while most of the Seyfert 2s are classified correctly as LINERs. The majority of LINER objects are correctly identified, while the majority of HII optically classified ULIRGs are spread across HII and LINER groups. Both the LINER and the HII classifications lie in similar areas of PCA space, and discrete classification for objects in this region may not be completely appropriate as many ULIRGs will show

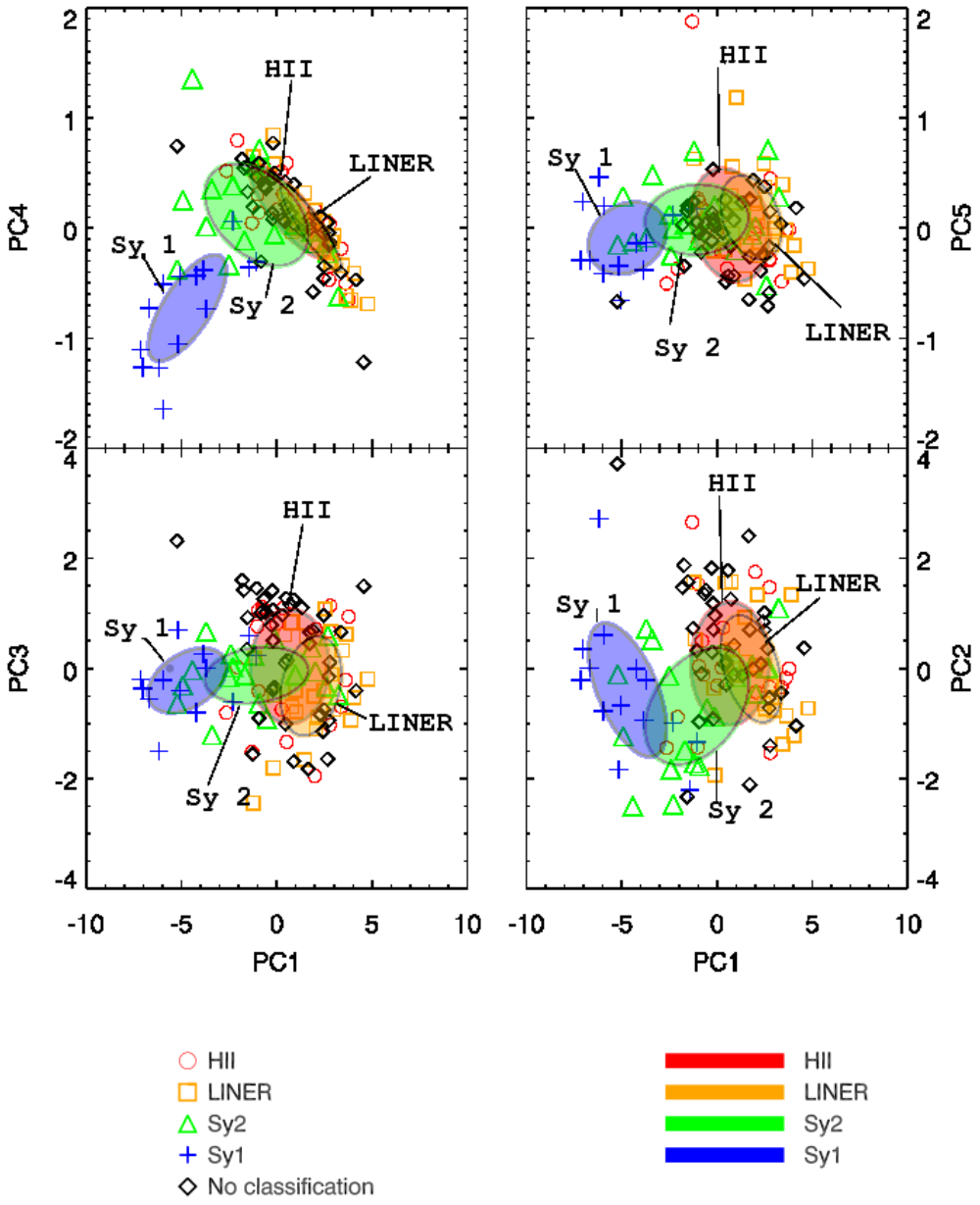


Figure 2.11: Four out of the possible 10 2D projections for our PCA space with the one sigma contours for the gaussian mixtures based classifications. Optically classified Seyfert 1 objects are marked by crosses, Seyfert 2 by triangles, LINERs by squares and HII classified objects with open circles. Those objects without optical classification are marked by diamonds.

| | | Gaussian classification | | | |
|---|---|---|---|---|---|
| | | HII | LINER | Sy 2 | Sy 1 |
| Optical | HII | 40% | 45% | 15% | 0% |
| | LINER | 11% | 85% | 4% | 0% |
| | Sy 2 | 5% | 26% | 51% | 18% |
| | Sy 1 | 0% | 0% | 6% | 94% |
| Not classified | | 42% | 46% | 12% | 0% |

Table 2.3: The percentage of objects in the four classifications as a function of their original classification. Not classified refers to those objects without an optical classification.

signs of both. Overall our 5D Gaussian classification scheme works well in associating regions in PCA space with type of object and is a powerful tool in objectively classifying objects.

We have used our classification scheme to classify the 41 ULIRGs with no optical classification. The percentages can be seen in Table 2.3. We find the majority are HII and LINER objects while 12% are classified as Seyfert 2 like objects. None of the objects appear to be Seyfert 1, suggesting optical classification of Seyfert 1 objects is complete. We now make use of our 5D Gaussian classification scheme by creating average spectra for our four classifications using all 119 ULIRGs. Before averaging the spectra, each spectrum is normalised so that the mean flux over the whole wavelength range is unity. The resulting four average templates can be seen in Figure 2.12. As expected, the HII and LINER templates are similar, whilst Seyfert templates have very little PAH emission.

## 2.7 Conclusions

We have shown that five principal components are needed to describe most of the variation in the 119 local ULIRG sample and are more successful than a full $\chi^2$ fitting by radiative transfer models. We have examined what the radiative transfer models are missing. The fits provided by radiative transfer models appear to need a cold dust component and have difficulty in modelling the strength of strong PAH emission lines.

We have used a combination of the Siebenmorgen & Krügel (2007) starburst models and Efstathiou & Rowan-Robinson (1995) AGN torus templates to investig-

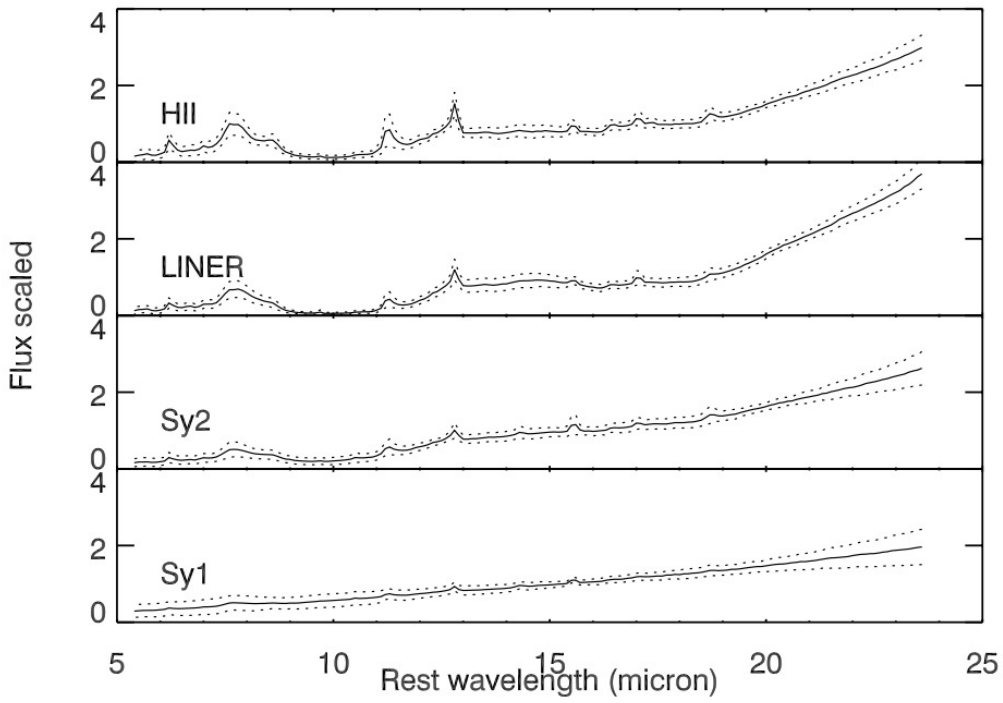


Figure 2.12: The average spectra for the four classifications, HII (29 objects), LINER (54 objects), Seyfert 2 (20 objects) and Seyfert 1 (16 objects). The dotted lines represent the one sigma dispersion in each classification.

ate what physical parameters are behind the components. We have examined how best fit model parameters are related to PC contribution. Overall, our conclusions are consistent with those reached in Wang et al. (2011). Contributions from PC1 appear to indicate the viewing angle of AGN with negative contributions associated with face on AGN and positive for obscured AGN. PC3 appears to be the best indicator of whether it is the AGN or starburst that is the prevailing power source.

The PCs consider a large part of the mid-infrared spectrum and are therefore less likely to be affected by problems associated with diagnostics based on single spectral features such as the PAH emission lines, where measuring line strength can be difficult. We suggest the five PCs would be useful as empirical templates for ULIRG spectra in the IRS public database (Lebouteiller et al., 2011).

We also introduce a new Gaussian mixtures classification scheme based on location in the five dimensional PCA space and trained via optical classifications. Objects can be classified as either Seyfert 1, Seyfert 2, LINER or HII-like. We note that any ULIRG with IRS spectra (in the relevant wavelength range) can be decomposed onto the PCs, and the position in PCA space can be used to classify the object.

We have used our classification scheme to provide a set of average spectra for the four groups. We make these, the five PCs and code to classify objects available at: `http://www.phys.susx.ac.uk/~pdh21/PCA/`

## Acknowledgements

We acknowledge support from the Science and Technology Facilities Council [grant numbers ST/F006977/1, ST/I000976/1, PP/E005306/1]. This work is based on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. We thank the referee for the very helpful comments which led to improvements in this paper.

# Bibliography

Alexander D. M., Efstathiou A., Hough J. H., Aitken D. K., Lutz D., Roche P. F., Sturm E. (1999). Observations and a model for the infrared continuum of centaurus a. *Monthly Notices of the Royal Astronomical Society*, 310:78. Cited on 33

Armus L., et al. (2007). Observations of ultraluminous infrared galaxies with the infrared spectrograph on the spitzer space telescope. ii. the iras bright galaxy sample. *The Astrophysical Journal*, 656:148. Cited on 32

Barger A. J., Cowie L. L., Sanders D. B., Fulton E., Taniguchi Y., Sato Y., Kawara K., Okuda H. (1998). Submillimetre-wavelength detection of dusty star-forming galaxies at high redshift. *Nature*, 394:248. Cited on 30

Bromley B. C., Press W. H., Lin H., Kirshner R. P. (1998). Spectral classification and luminosity function of galaxies in the las campanas redshift survey. *The Astrophysical Journal*, 505:25. Cited on 31

Connolly A. J., Szalay A. S., Bershady M. A., Kinney A. L., Calzetti D. (1995). Spectral classification of galaxies: an orthogonal approach. *Astronomical Journal v.110*, 110:1071. Cited on 31

Davoodi P., et al. (2006). Parametric modeling of the 3.6-8 $\mu$m color distributions of galaxies in the swire survey. *The Astronomical Journal*, 132:1818. Cited on 48, 57

Eales S., Lilly S., Webb T., Dunne L., Gear W., Clements D., Yun M. (2000). The canada-uk deep submillimeter survey. iv. the survey of the 14 hour field. *The Astronomical Journal*, 120:2244. Cited on 30

Efstathiou A., Rowan-Robinson M. (1995). Dusty discs in active galactic nuclei. MNRAS, 273:649–661. Cited on 32, 33, 50

Efstathiou A., Rowan-Robinson M., Siebenmorgen R. (2000). Massive star formation in galaxies: radiative transfer models of the uv to millimetre emission of starburst galaxies. MNRAS, 313:734–744. Cited on 33

Efstathiou A., Siebenmorgen R. (2005). Iso observations and models of galaxies with hidden broad line regions. *Astronomy and Astrophysics*, 439:85. Cited on 34

Efstathiou A. (2006). A model for the infrared emission of fsc 10214+4724. *Monthly Notices of the Royal Astronomical Society: Letters*, 371:L70. Cited on 33

Efstathiou A., Siebenmorgen R. (2009). Starburst and cirrus models for submillimeter galaxies. *Astronomy and Astrophysics*, 502:541. Cited on 33

Farrah D., Serjeant S., Efstathiou A., Rowan-Robinson M., Verma A. (2002). Submillimetre observations of hyperluminous infrared galaxies. *Monthly Notice of the Royal Astronomical Society*, 335:1163. Cited on 33

Farrah D., Afonso J., Efstathiou A., Rowan-Robinson M., Fox M., Clements D. (2003). Starburst and agn activity in ultraluminous infrared galaxies. *Monthly Notice of the Royal Astronomical Society*, 343:585. Cited on 33

Farrah D., et al. (2007). High-resolution mid-infrared spectroscopy of ultraluminous infrared galaxies. *The Astrophysical Journal*, 667:149. Cited on 31, 32

Farrah D., et al. (2008). The nature of star formation in distant ultraluminous infrared galaxies selected in a remarkably narrow redshift range. *The Astrophysical Journal*, 677:957. Cited on 31

Farrah D., et al. (2009). An evolutionary paradigm for dusty active galaxies at low redshift. *The Astrophysical Journal*, 700:395. Cited on 31

Farrah D., et al. (2012). Direct evidence for termination of obscured star formation by radiatively driven outflows in reddened qsos. *The Astrophysical Journal*, 745:178. Cited on 34

Fox M. J., et al. (2002). The scuba 8-mjy survey - ii. multiwavelength analysis of bright submillimetre sources. *Monthly Notices of the Royal Astronomical Society*, 331:839. Cited on 30

Genzel R., et al. (1998). What powers ultraluminous iras galaxies? *Astrophysical Journal v.498*, 498:579. Cited on 31

Houck J. R., et al. (2004). The infrared spectrograph (irs) on the spitzer space telescope. *The Astrophysical Journal Supplement Series*, 154:18. Cited on 30

Hughes D. H., et al. (1998). High-redshift star formation in the hubble deep field revealed by a submillimetre-wavelength survey. *Nature*, 394:241. Cited on 30

Imanishi M., Dudley C. C., Maiolino R., Maloney P. R., Nakagawa T., Risaliti G. (2007). A spitzer irs low-resolution spectroscopic search for buried agns in nearby ultraluminous infrared galaxies: A constraint on geometry between energy sources and dust. *The Astrophysical Journal Supplement Series*, 171:72. Cited on 31, 32

Kessler M. F., et al. (1996). The infrared space observatory (iso) mission. *Astron. Astrophys.*, 315:L27. Cited on 30

Kruegel E. (2003). *The physics of interstellar dust*. IOP Publishing Ltd. Cited on 42

Le Floc'h E., et al. (2005). Infrared luminosity functions from the chandra deep field south : the spitzer view on the history of dusty star formation at 0¡z¡1. *arXiv*, astro-ph, astro-ph/0506462v1. Cited on 30

Lebouteiller V., Barry D. J., Spoon H. W. W., Bernard-Salas J., Sloan G. C., Houck J. R., Weedman D. W. (2011). Cassis: The cornell atlas of spitzer/infrared spectrograph sources. *The Astrophysical Journal Supplement*, 196:8. Cited on 52

Moorwood A. F. M. (1986). 3.28 micron feature and continuum emission in galaxy nuclei. *Astronomy and Astrophysics (ISSN 0004-6361)*, 166:4. Cited on 31

Nardini E., Risaliti G., Salvati M., Sani E., Watabe Y., Marconi A., Maiolino R. (2009). Exploring the active galactic nucleus and starburst content of local ultraluminous infrared galaxies through 5-8 m spectroscopy. *Monthly Notices of the Royal Astronomical Society*, 399:1373. Cited on 31

Rieke G. H., Low F. J. (1972). Infrared photometry of extragalactic sources. *Astrophysical Journal*, 176:L95. A&AA ID. AAA008.158.057. Cited on 30

Rigopoulou D., Spoon H. W. W., Genzel R., Lutz D., Moorwood A. F. M., Tran Q. D. (1999). A large mid-infrared spectroscopic and near-infrared imaging survey of ultraluminous infrared galaxies: Their nature and evolution. *The Astronomical Journal*, 118:2625. Cited on 31

Roche P. F., Aitken D. K., Smith C. H., Ward M. J. (1991). An atlas of mid-infrared spectra of galaxy nuclei. *Royal Astronomical Society*, 248:606. Cited on 31

Rowan-Robinson M., et al. (1997). Observations of the hubble deep field with the infrared space observatory - v. spectral energy distributions, starburst models and star formation history. *Monthly Notices of the Royal Astronomical Society*, 289:490. Cited on 30

Ruiz M., Efstathiou A., Alexander D. M., Hough J. (2001). Constraints on the nuclear emission of the circinus galaxy: the torus. *Monthly Notices of the Royal Astronomical Society*, 325:995. Cited on 34

Sanders D. B. (1999). A new view of galaxy evolution from submillimeter surveys with scuba. *Astrophysics and Space Science*, 269:381. Cited on 30

Siebenmorgen R., Krügel E. (2007). Dust in starburst nuclei and ulirgs. sed models for observers. A&A, 461:445–453. Cited on 32, 33, 42, 50

Soifer B. T., et al. (1984). Infrared galaxies in the iras minisurvey. *Astrophysical Journal*, 278:L71. Cited on 30

Spoon H. W. W., Marshall J. A., Houck J. R., Elitzur M., Hao L., Armus L., Brandl B. R., Charmandaris V. (2007). Mid-infrared galaxy classification based on silicate obscuration and pah equivalent width. *The Astrophysical Journal*, 654:L49. Cited on 31, 32

Verma A., Rowan-Robinson M., McMahon R., Efstathiou A. (2002). Observations of hyperluminous infrared galaxies with the infrared space observatory: implications for the origin of their extreme luminosities. *Monthly Notice of the Royal Astronomical Society*, 335:574. Cited on 33

Wang L., Farrah D., Connolly B., Connolly N., LeBouteiller V., Oliver S., Spoon H. (2011). Principal component analysis of the spitzer irs spectra of ultraluminous infrared galaxies. *Monthly Notices of the Royal Astronomical Society*, 411:1809. Cited on 29, 31, 32, 34, 35, 36, 37, 38, 42, 48, 52

Werner M. W., et al. (2004). The spitzer space telescope mission. *The Astrophysical Journal Supplement Series*, 154:1. Cited on 30

## 2.8 (Appendix A) Gaussian Mixtures Modelling

When using GMM as part of a classifier, the available data $\chi_{opt.}$ from each optical classification is modelled with a combination of multivariate Gaussians. For our classifier, since the number of data points is small, we model each optical classification with one multivariate Gaussian. The probability density function for a multivariate Gaussian is:

$$p(\mathbf{x}|\mathbf{m},\mathbf{S}) = \frac{1}{\sqrt{\det(2\pi\mathbf{S})}}\exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{x}-\mathbf{m})\} \tag{2.2}$$

where $\mathbf{x}$ is a point in the $D$ dimensional space, $\mathbf{m}$ is the mean and $\mathbf{S}$ is the covariance matrix.

By restricting the number of Gaussians to one, for each optical classification $c$, we calculate the mean and covariance of $\chi_c$ to get $\mathbf{m}^c$ and $\mathbf{S}^c$, which when combined with equation 2.2 gives $p(\mathbf{x}|c,\chi_c)$.

For any point in our N dimensional space, $x^*$, the posterior probability of belonging to a particular class $c$, is given by:

$$p(c|\mathbf{x}^*,\chi) \propto p(\mathbf{x}^*|c,\chi_c)p(c) \tag{2.3}$$

Where $p(c)$ is the is the prior class probability. For our classifier, we assign an equal prior probability of belonging to each class. Using the highest probability density provides the optimal (maximum likelihood) classification. However, since the PDFs overlap, this will not provide the best classification for the population statistics. We therefore take the same approach as Davoodi et al. (2006) and randomly assign each galaxy to a class, with probability proportional to the PDF values at the galaxies position in PCA space.

# Chapter 3

# Learning the fundamental mid-infrared spectral components of galaxies with non-negative matrix factorization

*Peter Hurley, Seb Oliver, Duncan Farrah, Vianney Lebouteiller, Henrik Spoon*

## 3.1 Abstract

The mid-infrared (MIR) spectra observed with the *Spitzer* Infrared Spectrograph (IRS) provide a valuable dataset for untangling the physical processes and conditions within galaxies.

This paper presents the first attempt to blindly learn fundamental spectral components of MIR galaxy spectra, using non-negative matrix factorisation (NMF). NMF is a recently developed multivariate technique shown to be successful in blind source separation problems. Unlike the more popular multivariate analysis technique, principal component analysis, NMF imposes the condition that weights and spectral components are non-negative. This more closely resembles the physical process of emission in the mid-infrared, resulting in physically intuitive components. By applying NMF to galaxy spectra in the Cornell Atlas of Spitzer/IRS sources (CASSIS), we find similar components amongst different NMF sets. These similar com-

ponents include two for AGN emission and one for star formation. The first AGN component is dominated by fine structure emission lines and hot dust, the second by broad silicate emission at 10 and 18 $\mu$m. The star formation component contains all the PAH features and molecular hydrogen lines. Other components include rising continuums at longer wavelengths, indicative of colder grey-body dust emission. We show an NMF set with seven components can reconstruct the general spectral shape of a wide variety of objects, though struggle to fit the varying strength of emission lines. We also show that the seven components can be used to separate out different types of objects. We model this separation with Gaussian Mixtures modelling and use the result to provide a classification tool.

We also show the NMF components can be used to separate out the emission from AGN and star formation regions and define a new star formation/AGN diagnostic which is consistent with all mid-infrared diagnostics already in use but has the advantage that it can be applied to mid-infrared spectra with low signal to noise or with limited spectral range. The 7 NMF components and code for classification are made public on arxiv and are available at: `https://github.com/pdh21/NMF_software/`.

## 3.2 Introduction

Spectra of the integrated mid-infrared (MIR) emission from galaxies contain a wealth of diagnostics that probe the origin of their MIR luminosity. For example, the main polycyclic aromatic hydrocarbons (PAHs) emission features found at 6.2, 7.7, 8.6, 11.3 and 12.7 $\mu$m are strong in objects where star formation activity contributes significantly to the mid-IR luminosity (Genzel et al., 1998; Laurent et al., 2000). The PAH features are either weak or absent for objects dominated by an active galactic nucleus (AGN) while emission lines with a high ionisation potential, for example the Neon fine structure line [Ne V] 14.3 $\mu$m, tend to be strong in the presence of an AGN (Genzel et al., 1998; Sturm et al., 2000). Ratios of other fine structure lines such as [Ne III] $\mu$m 15.56 / [Ne II] 12.81 $\mu$m versus [S III] 33.48 $\mu$m/[Si II] 34.82 $\mu$m have been shown to diagnose power source (Dale et al., 2006) as has the shape of the underlying mid-infrared dust continuum. (Brandl et al., 2006).

Observations from the Infrared Space Observatory (ISO: (Kessler et al., 1996)),

and the Infrared Spectrograph (IRS; (Houck et al., 2004)) on the Spitzer Space Telescope (Werner et al., 2004) allowed the MIR spectral features to be used as diagnostics of star formation and AGN activity. Combinations of the PAH emission lines, high to low excitation mid-infrared emission lines, silicate features and continuum measurements have been used as diagnostics for characterising the power source behind Ultraluminous Infrared Galaxies (ULIRGs) (Genzel et al., 1998; Rigopoulou et al., 1999; Spoon et al., 2007; Farrah et al., 2007, 2008, 2009; Petric et al., 2010).

However, diagnostics based on specific emission and absorption lines only focus on small parts of the spectrum, disregarding the information contained in the rest of the mid-infrared region. They can also be ambiguous. Dust and gas require ionising radiation to emit in the mid-IR, the source of the radiation is not important. For example, hot OB stars or an accretion disk around a supermassive black hole can both produce the [OIV] 25.9 $\mu$m emission line, as well as shocks (e.g. Lutz et al., 1998). The line ratios of fine structure lines can also be affected by the geometry of the emitting region and the age of a starburst, while the metallicity can affect PAH emission strength (e.g. Thornley et al., 2000; Engelbracht et al., 2005; Madden et al., 2006; Wu et al., 2006; Farrah et al., 2007). As a result, different diagnostics can give conflicting estimates for the contribution from star formation and/or AGN (e.g. Armus et al., 2007; Veilleux et al., 2009).

Separation of spectral features from continuum and the mixing of neighbouring spectral features can also be problematic. For example, measurement of the 9.7 $\mu$m silicate feature requires different methods depending on the strength of the 8.6 and 11.2 $\mu$m PAH emission lines (Spoon et al., 2007).

An alternative method for identifying the power source is to decompose the spectra with AGN and starburst spectral templates. These templates tend to be a spectrum from a specific object (e.g. M82) or a mean spectrum of a number of similar object types. Pope et al. (2008) use a combination of the M82 spectrum, average spectral template of starburst galaxies (Brandl et al., 2006) and a power law to decompose the IRS spectra of 13 high redshift submillimeter galaxies. Valiante et al. (2009) fit IRS spectra across the range 5.5-6.85 $\mu$m with a combination of the M82 spectrum and a linear approximation for the AGN continuum. Alonso-Herrero et al. (2011) use the (Brandl et al., 2006) starburst template and CLUMPY radiative

transfer models for AGN to decompose the IRS spectra of 53 LIRGs into starburst and AGN components. Using average starburst templates is both simplistic and problematic. Prior theoretical prejudices drive the choice for what objects are used for the average templates, and they may be contaminated by AGN emission. The same is true for AGN average spectral templates.

With the public release of all low resolution Spitzer/IRS spectra by the Cornell Atlas of Spitzer/IRS sources (CASSIS)(Lebouteiller et al., 2011)*, we are now in a better position to investigate the role played by star formation and AGN with more sophisticated techniques. In this paper we use a multivariate analysis technique to blindly learn the fundamental MIR spectral components, which we interpret as different physical environments within galaxies. Learning the MIR spectral shape of physical environments, allows the whole MIR wavelength range to be used as a diagnostic. The spectral components also provide an alternative to average spectral templates.

A subclass of multivariate analysis techniques include matrix factorisation algorithms. The techniques are often associated with pattern recognition and blind source separation (Lee & Seung, 2001). Algebraically, the algorithms approximate a data matrix by two simpler matrices: a weight matrix and component matrix. Common factorisation techniques include Singular Value Decomposition, Principal Component Analysis and Independent Component Analysis. The different techniques use different assumptions to carry out the factorisation, resulting in different weights and components. As multivariate datasets of spectra have become more prevalent, techniques such as Principal Component Analysis (PCA) have been applied to astronomical problems. PCA has already been used for spectral classification of optical galaxies (e.g. Connolly et al., 1995; Bromley et al., 1998; Taghizadeh-Popp et al., 2012). PCA has also been successfully applied to the IRS spectra of local ULIRGs (Wang et al., 2011; Hurley et al., 2012).

The weights and spectral templates derived with PCA can be both positive and negative. Spectral reconstruction involves both addition and cancellation of spectral features. As a result, the PCA templates are inherently difficult to interpret physically.

A relatively new matrix factorisation technique, Non-negative matrix factorisa-

*The Cornell Atlas of Spitzer/IRS Sources (CASSIS) is a product of the Infrared Science Center at Cornell University, supported by NASA and JPL.

tion (NMF; Lee & Seung (1999)) can be thought of as PCA but with non-negative constraints on weights and templates. The constraints make reconstruction a purely additive process which more closely resembles emission in the mid-infrared. The first application of NMF to astronomy was carried out by Blanton & Roweis (2007) who adopted the Lee & Seung (2001) NMF algorithms and applied it to optical spectra and photometry. It has also been used as a blind source separation algorithm on the IRS spectra of galactic photo-dissociation regions (Berné et al., 2007; Rosenberg et al., 2011).

This paper presents the first NMF analysis on mid-infrared galaxy spectra. We use spectra from the recently released Cornell Atlas of Spitzer/IRS sources (CASSIS) (Lebouteiller et al., 2011). Our paper provides the first large scale statistical analysis of the IRS spectra to date using the NMF algorithm. Section 3.3 describes the CASSIS database and data reduction. In Section 3.4, we describe the suitability of matrix factorisation to IRS spectra, and give details on the NMF algorithm. In section 3.5 we present our results and in Section 3.6 our conclusions. We assume a spatially flat cosmology with $H_0 = 70 \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}, \Omega = 1$, and $\Omega_m = 0.3$.

## 3.3 The Data

### 3.3.1 CASSIS

We use spectra from the Cornell Atlas of Spitzer/IRS sources (CASSIS) (Lebouteiller et al., 2011). The atlas contains sources observed in low resolution mode with the Infrared Spectrograph (IRS;Houck et al. (2004)) on board the *Spitzer Space Telescope* (Werner et al., 2004). IRS low resolution mode observations were made using two low-resolution modules, ShortLow and LongLow (hereafter SL and LL), covering 5.2-14.5 and 14.0-38.0 $\mu$m respectively. The modules also had a resolving power of $R \approx 60 - 120$ ($\approx 75\%$ of the observations) and an aperture size of $3.7 \times 57''$ for SL and $10.7 \times 168''$ for LL. The observations in the CASSIS database are first processed with the Basic Calibrated Data (BCD) pipeline from the *Spitzer* Science pipeline (release S18.7.0.) and produces BCD frames. This removes electronic and optical artefacts. The BCD images are then processed using the CASSIS pipeline which carries out image cleaning, background subtraction, and spectral extraction. The pipeline algorithm is both automatic and flexible enough to handle different

observations, from barely detected sources to bright sources and from point-like to somewhat extended sources.

### 3.3.2 Sample

The current version of CASSIS (version 4) contains 11304 distinct sources. 2118 of those distinct sources have known spectroscopic redshifts taken from NASA/IPAC Extragalactic Database (NED*). We make the additional redshift cut ($0.01 < z < 0.2$). The lower limit prevents contamination from Galactic and local group sources while the upper limit ensures we sample approximately the same wavelength range for each object. The redshift cut gives us a sample size of 893. We note that the redshifts within CASSIS, have been collected heterogeneously, biasing our sample by the parent redshift surveys. Because objects in our sample are at low redshift and span many programs, they likely span most or all IR luminous object types in the local Universe. Therefore, while a small degree of bias is inevitable, we do not consider that it is significant enough to significantly affect our results. We also only use objects with both SL and LL data. This reduces our sample size down to 729 objects. The redshift distribution for the 729 objects can be seen in Figure 3.1.

### 3.3.3 Stitching

Observations using data from both SL and LL spectral modules can suffer from mismatching due to telescope pointing inaccuracy or if a source is extended in SL and not in LL. The mismatching causes the spectra from one of the modules (normally the SL) to have lower flux calibration than the other. Correcting the mismatch is inherently difficult as the data from the overlap between the two modules can suffer from the '14 micron teardrop' (see IRS instrument handbook, †), leaving a small gap at around 13-14 $\mu$m.

We correct for the mismatch using a simplified version of our NMF technique. For the first step, we generated two sets of templates, one using SL data and the other using LL data. The distribution in redshift causes the mismatch region to occur at different rest frame wavelengths for different objects. This ensures at least one template set covered the mismatch region for each object. We then fitted the template set to a region of width 7 $\mu$m, centred on the mismatch area. Wavelength

*http://nedwww.ipac.caltech.edu/

†http://irsa.ipac.caltech.edu/data/SPITZER/docs/irs/

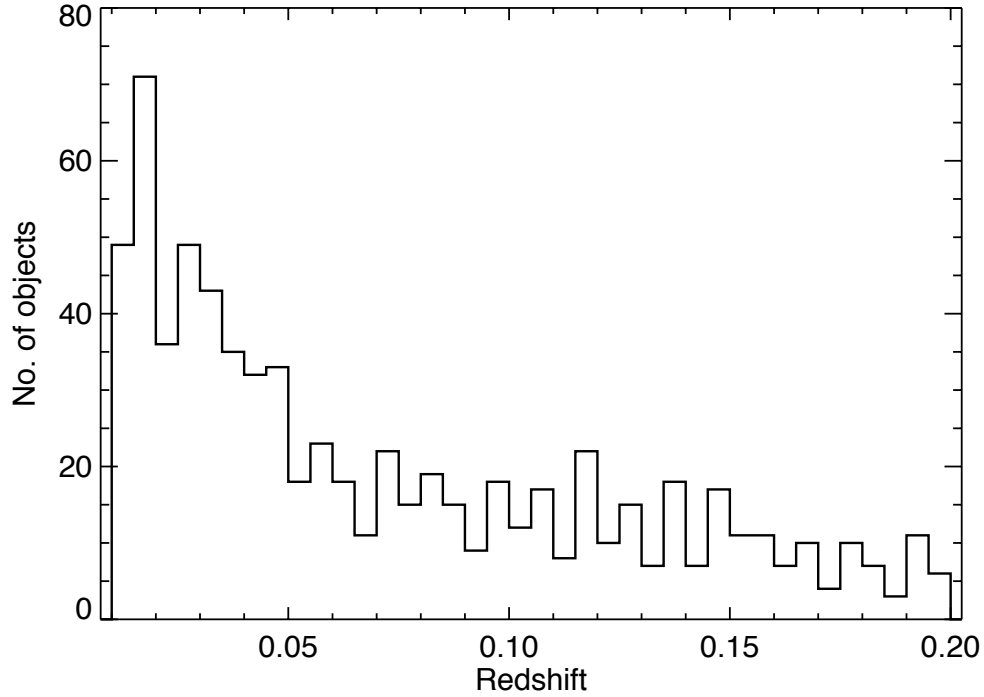


Figure 3.1: The redshift distribution for the sample selection we apply the NMF algorithm to.

points associated with PAH and Neon emission lines were removed to prevent strong line strengths from distorting the fits. We carry out the fit for different scalings applied to the SL data. The scaling factor value that gives the lowest $\chi^2$ is chosen as the scaling correction. Having stitched the spectra using both SL and LL template sets, we then generated our initial NMF sets for the entire spectral range. We then re-stitch the spectra with the new NMF set. Additional spectra used for analysis in this paper are also stitched with our final $NMF_7$ set, introduced in section 3.5.

### 3.3.4 Normalisation

The NMF analysis requires all spectra to be normalised to a standard value to prevent sources with higher flux, biasing the algorithm. We normalise all the spectra by the average flux across the restframe wavelength range of $7 - 20\mu$m. We choose this range as it is common to all sources with both SL and LL data.

## 3.4 Matrix Factorisation

Analysis of spectra from Spitzer's IRS has tended to be done using diagnostics based on only a few of the specific features (e.g. Sajina et al., 2007; Pope et al., 2008; Alonso-Herrero et al., 2012). For example, Spoon et al. (2007) introduced a classification scheme based on the 6.2$\mu$m PAH line and 10$\mu$m silicate feature. Quantifying the contribution from star formation and AGN has also been carried out using fine structure lines, for example the [OIV]/[NeII] and [NeV]/[NeII] line ratios versus the 6.2$\mu$m PAH equivalent width. (e.g. Armus et al., 2007; Petric et al., 2010)

In essence, line diagnostic analyses are carrying out a crude compression by using only small parts of the spectrum to describe each object (e.g. the 6.2$\mu$m feature). Matrix factorisation techniques provide an alternative approach to compression by transforming data from wavelength space to one that better captures the variance in the dataset. As a result, classification or quantification of properties such as star formation is carried out considering a greater wavelength range.

Algebraically, matrix factorisations find a linear approximation to a data matrix X such that $\mathrm{X} \approx \mathrm{WH}$, or:

$$\mathrm{X}_{i\mu} \approx (\mathrm{WH})_{i\mu} = \sum_{a=1}^{r} \mathrm{W}_{ia}\mathrm{H}_{a\mu} \tag{3.1}$$

Where, $i$ is object index, $\mu$ is wavelength and $a$ is component index. The matrix H can be thought of as a set of $r$ components that represent latent structure explicit in the dataset, and W are a set of weighting coefficients. Each object in the dataset can now be approximated by a linear combination of the derived components, H.

Different matrix factorisation techniques use different assumptions to carry out the approximation. Independent component analysis (ICA) assumes the derived components (H) are independent. Principal component analysis (PCA) models the dataset as a multivariate Gaussian distribution in wavelength space and finds the orthogonal components of the Gaussian. Non-Negative Matrix Factorisation (NMF) assumes the data, weights and components are all non-negative, but makes no assumption on the distribution of the data or correlation between derived components.

### 3.4.1 Matrix factorisation of spectra

By applying linear matrix factorisation techniques to the mid-infrared spectra of galaxies, we are assuming mid-infrared spectra of galaxies, $F(\lambda)$, can be modelled as a linear combination of components. Ideally the components would relate to physical regions, for example a star forming region ($T_{SF}$), an active galactic nuclei torus ($T_{AGN}$), a molecular cloud ($T_{MC}$) or diffuse dust component ($T_C$). A spectrum for a galaxy would then simply be:

$$F(\lambda) = a \cdot T_{SF}(\lambda) + b \cdot T_{AGN}(\lambda) + c \cdot T_{MC}(\lambda) + d \cdot T_C(\lambda) \tag{3.2}$$

Where, $a, b, c$ and $d$ are the relative weights for each component.

For the above model, ICA is not suitable as the components are unlikely to be independent, for example AGN and star formation are believed to be triggered by similar mechanisms such as mergers (e.g. Sanders et al., 1988), and are likely to be connected through feedback processes (e.g. Farrah et al., 2012; Rovilos et al., 2012).

PCA has already been applied to the mid-infrared spectra of ULIRGs (e.g. Wang et al., 2011; Hurley et al., 2012). Algebraically, PCA calculates the eigenvectors of the covariance matrix. For spectra, the principal components represent the principal variations from a mean spectral template. The components are therefore allowed to have features which are positive and negative, and are also allowed to have a negative weighting when fitting objects. The freedom to be both positive and negative does not mimic the process of emission in the MIR, resulting in components that are inherently difficult to interpret. By their nature, the principal components have a statistical rather than physical interpretation. Therefore, although PCA can successfully reduce dimensionality of spectra for classification from known objects, it is not suitable for our model.

The non-negative constraint of NMF more closely reflects the physical process of emission in the mid-infrared, which does not suffer from the same problems of absorption as other spectral ranges. As a result the NMF generated templates are more physically intuitive.

NMF is therefore the most applicable matrix factorisation routine for our linear interpretation of galaxy emission. However, the situation is complicated by dust extinction. This introduces a non-linearity to the problem since extinction is multiplicative and exponential.

$$F(\lambda) = (a \cdot T_{SF}(\lambda) + b \cdot T_{AGN}(\lambda) + c \cdot T_{MC}(\lambda) + d \cdot T_C(\lambda))e^{-f \cdot \tau(\lambda)} \tag{3.3}$$

Where $f$ is the weight associated with extinction and $\tau(\lambda)$ can either be known or unknown.

We can take the model one step further by allowing extinction to vary across all four components:

$$F(\lambda) = a \cdot T_{SF}(\lambda)e^{-f \cdot \tau(\lambda)} + b \cdot T_{AGN}(\lambda)e^{-g \cdot \tau(\lambda)} + c \cdot T_{MC}(\lambda)e^{-h \cdot \tau(\lambda)} + d \cdot T_C(\lambda)e^{-i \cdot \tau(\lambda)} \tag{3.4}$$

The weights for the extinction are $f, g, h$ and $i$.

We have explored the suitability of non-linear kernel based matrix factorisation algorithms (e.g. Zafeiriou & Petrou, 2010; Pan et al., 2011) and found they are not suited for the non-linear behaviour described in equations 3.3 and 3.4. We discuss why in Appendix A. Current algorithms therefore restrict us to describe mid-infrared galaxy spectra as a set of linear components (e.g. equation 3.2) and NMF is the most appropriate matrix factorisation technique.

The first application of NMF in astronomy was carried out by Blanton & Roweis (2007) who updated the popular NMF multiplicative algorithm from Lee & Seung (2001) to include uncertainties and for heterogeneous datasets (e.g. optical spectra and photometric observations of galaxies at different redshifts). They also restricted the space of possible spectra to those predicted from high resolution stellar population synthesis models. We use the NMF algorithm from Blanton & Roweis (2007) to identify and learn the mid-infrared sources that are common to galaxies in the CASSIS database. Unlike Blanton & Roweis (2007), we do not use any models as a prior for shape of the components, we use the algorithm to blindly learn the shape of our components.

### 3.4.2 NMF algorithm

As with PCA, the goal of NMF is to minimise a cost function. The most widely used is the squared approximation error described in Lee & Seung (2001):

$$\chi^2 = \sum_{i\mu} \left( \mathrm{X}_{i\mu} - \sum_a \mathrm{W}_{ia}\mathrm{H}_{a\mu} \right)^2 \tag{3.5}$$

Minimising equation 3.5 requires some sort of numerical technique to find local minima. Lee & Seung (2001) presented 'multiplicative update rules' for H and W. Upon each iteration, the rules are used to update H and W by a multiplicative factor whilst minimising equation 3.5. The algorithm implemented in Blanton & Roweis (2007) altered the original multiplicative update algorithm of Lee & Seung (2001) for nonuniform uncertainties ($\sigma$). The cost function then becomes the weighted squared approximation error:

$$\chi^2 = \sum_{i\mu} \left( \frac{\mathrm{X}_{i\mu} - \sum_a \mathrm{W}_{ia}\mathrm{H}_{a\mu}}{\sigma_{i\mu}} \right)^2 \tag{3.6}$$

Blanton & Roweis (2007) showed the multiplicative update rules for H and W are as follows:

$$\mathrm{W}_{ia} \leftarrow \mathrm{W}_{ia} \left( \sum_{\mu} \frac{\mathrm{X}_{i\mu}\mathrm{H}_{a\mu}}{\sigma_{i\mu}^2} \right) \left( \sum_{m\mu} \frac{\mathrm{W}_{im}\mathrm{H}_{m\mu}\mathrm{H}_{a\mu}}{\sigma_{i\mu}^2} \right)^{-1} \tag{3.7}$$

$$\mathrm{H}_{a\mu} \leftarrow \mathrm{H}_{a\mu} \left( \sum_{i} \frac{\mathrm{W}_{ia}\mathrm{X}_{i\mu}}{\sigma_{i\mu}^2} \right) \left( \sum_{mi} \frac{\mathrm{W}_{ia}\mathrm{W}_{im}\mathrm{H}_{m\mu}}{\sigma_{i\mu}^2} \right)^{-1} \tag{3.8}$$

The update rules in equations 3.7 and 3.8 are guaranteed to reduce the error, however the cost function in equation 3.6 is not necessarily convex therefore the algorithm may get stuck in a local minimum. We run the algorithm five times with different initial starting positions to check the solution is consistent.

Convergence can be evaluated by looking at the decrease in cost function across iterations and checking the solution has reached a minimum. In practise, we find 3000 iterations are enough for H and W to converge.

The number of components generated by NMF is a user input. Unlike PCA where the shape of the original components remain unchanged as more are added, the NMF components will not remain the same. We investigate the number of components required to constrain the data by generating 11 different NMF sets, containing from 3 up to 14 components. We define the following notation, $NMF_y^x$ to describe the $x$th component from an NMF set containing $y$ components.

### 3.4.3 Bayesian Evidence

To determine the minimum number of components that are justified by the data, one should calculate the Bayesian evidence ($E$).

$$E \equiv \int L(\theta)\pi(\theta)d\theta \tag{3.9}$$

The evidence can be thought of as the average likelihood, $L(\theta)$, over all of the prior, $\pi(\theta)$, parameter space, $d\theta$, of a given model and automatically implements Occam's razor, i.e. simpler models are preferred unless simplicity can be traded for greater explanatory power.

There are two ways in which one could calculate the Bayesian evidence for our setup. The first would be to calculate the evidence for the NMF algorithm, where the number of parameters is equal to the number of elements in both H and W. This approach would be the most appropriate if comparing the suitability of NMF to other matrix factorisation techniques, the integral however becomes highly multi-dimensional making the calculation numerically challenging. Alternatively, if NMF is the most appropriate algorithm to our problem, then we can assume that the components are correct. The number of parameters is then equal to the number of elements in W, i.e. the number of components.

We choose the later approach as we have already chosen NMF as the most appropriate algorithm to our problem and are not comparing alternative procedures.

We calculate the evidence by using the nested sampling routine, MULTINEST (Feroz et al., 2008) to re-fit the CASSIS sample with different NMF sets. MULTINEST is a Bayesian inference tool which calculates the evidence and produces posterior samples from distributions with (often an unknown number of) multiple modes and/or degeneracies between parameters. Nested Sampling (Skilling, 2004) is a Monte Carlo technique that randomly samples from the prior space, and zooms in on areas of higher likelihood during successive iterations.

We fit every galaxy with component sets $NMF_3$ to $NMF_{14}$ and their respective repeats. For every repeat, we calculate the median evidence of the sample. The main uncertainty on our evidence values comes from the difference in NMF sets across repeats (i.e. the convergence on slightly different local minima by the NMF algorithm). To quantify the uncertainty on our evidence values, we calculate the mean and standard deviation evidence values from the 5 repeats, as a function of number of components.

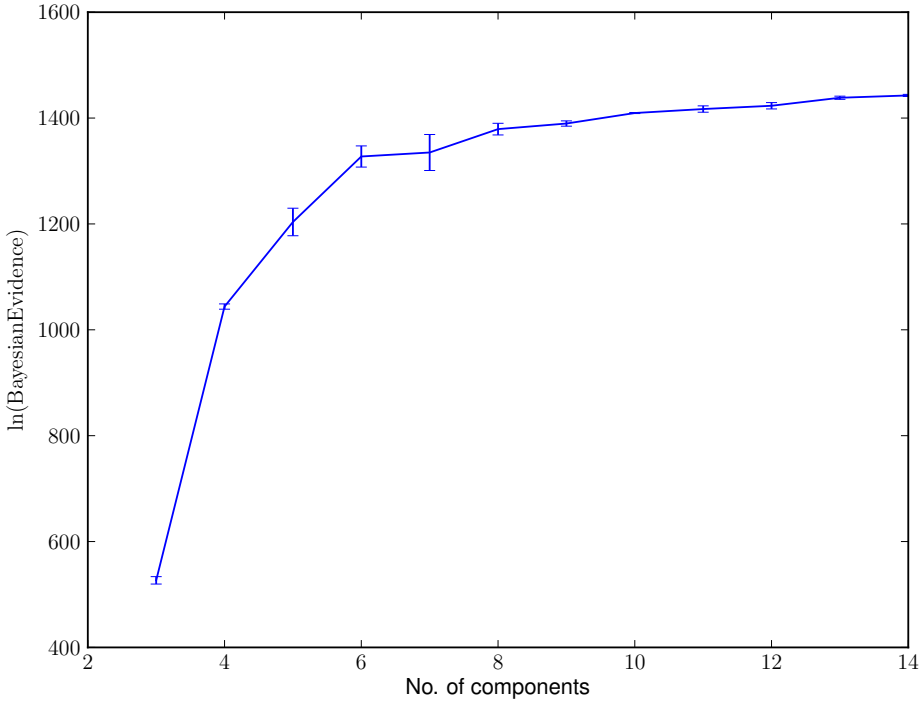


Figure 3.2: The Bayesian evidence as a function of number of components. For each NMF set, we run the algorithm 5 times and calculate the median evidence value of the entire galaxy sample. We plot the mean and standard deviation of the 5 repeats.

## 3.5 Results

### 3.5.1 Number of Components

As discussed in the previous section, we would like to quantify how many components are required by the data. Figure 3.2 shows the mean and standard deviation for the Bayesian evidence values from 5 repeats, as a function of number of components. The Bayesian evidence should start decreasing as the number of components exceeds the optimum number needed to constrain the data. We see no turnover, indicating there is not an obvious, optimum NMF set below 14 components. We note however a slight levelling of at 7 components before increasing again beyond 8.

We have also looked at the ratio of evidence values between consecutive NMF sets. The ratio, referred to as the Bayes factor ($K$), is used as a measure for a Bayesian version of classical hypothesis testing. We use the Jeffreys scale to interpret $K$. A value of $K < 1$ indicates the more complicated model is preferred, $K = 1 - 3$ as barely worth mentioning, $K = 3 - 10$ indicates substantial support for the simpler model, while $K = 10 - 30$ is strong, $K = 30 - 100$ is very strong, and $K > 100$ is considered decisive. Using the Jeffreys scale, we find more than 14 components are needed to reconstruct spectra within the uncertainties. However, we note that $K$ begins to level off after 6/7, indicating that although more complicated component sets are preferred, the gain in increasing the number of components is beginning to decrease.

Ideally, we would calculate the Bayesian evidence and Bayes factor beyond $NMF_{14}$. However, calculating evidence for highly multidimensional parameter spaces becomes computationally challenging. We have qualitatively examined NMF sets where number of components $> 14$. As an example, in Figure 3.15 we show the NMF components for $NMF_{30}$. Interpreting a many-component NMF set such as $NMF_{30}$ becomes challenging as signatures begin to separate out into several components, whose physical interpretation is not clear.

We also note that the Bayesian evidence calculation could be influenced by two fundamental factors. The first is the use of uncertainties associated with IRS spectra, which have often been underestimated below the observed variation between individual nod positions on the IRS, as described in Chapter 7 of the IRS Instrument Handbook *. As a result, our model selection may be too conservative. The

*http://irsa.ipac.caltech.edu/data/SPITZER/docs/irs/

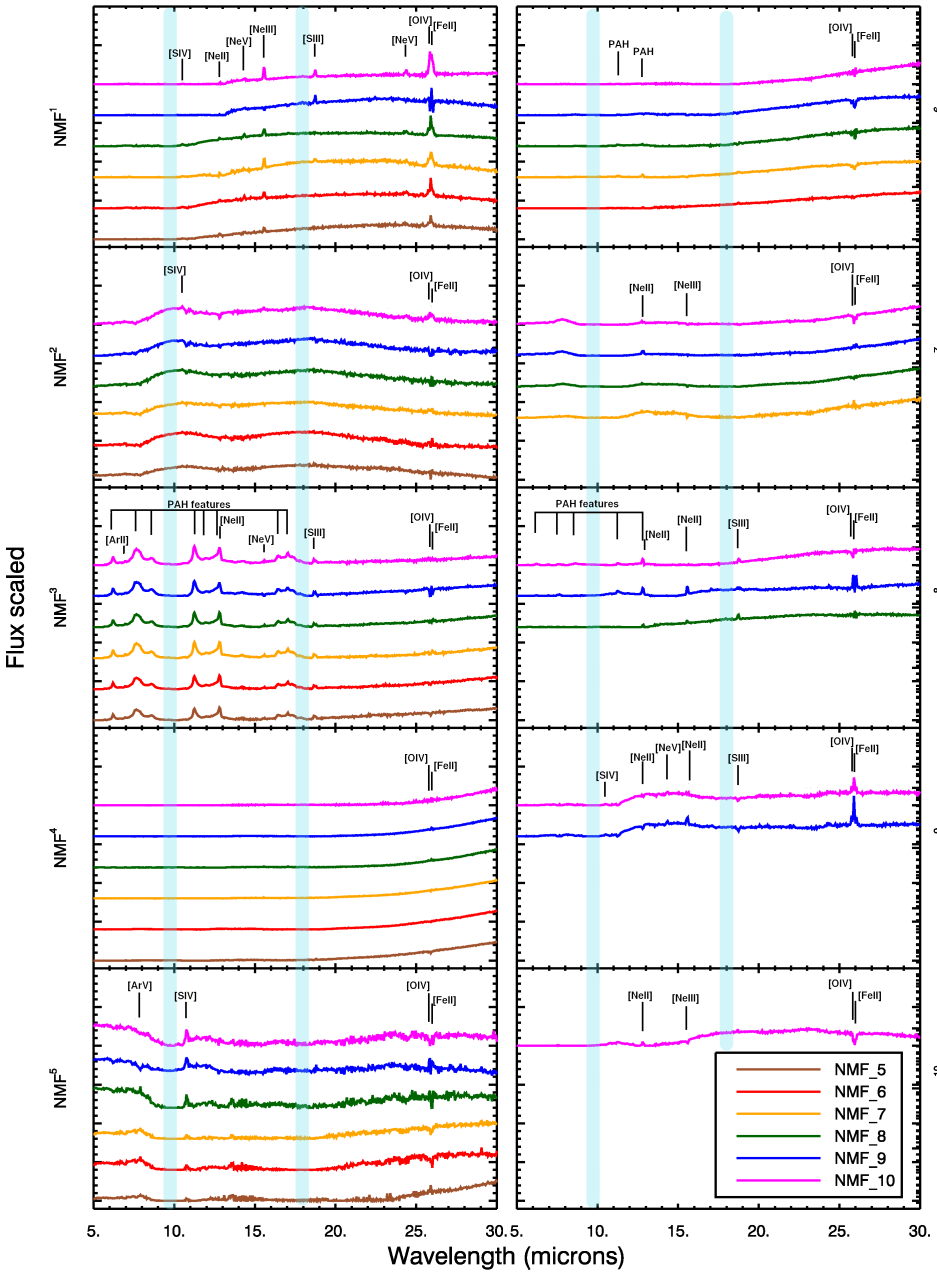


Figure 3.3: The derived NMF spectral components for sets $NMF_5$-$NMF_{10}$. Each NMF set is colour coded, with components ordered by similarity. For example, the five components of $NMF_5$ are the five brown spectra. Prominent spectral features in each component are also labelled and regions affected by broad silicate absorption and emission are highlighted in light blue.

other problem comes from the suitability of the NMF algorithm to the non-linear behaviour associated with extinction. We have carried out a simple simulation to show how extinction could be a factor in driving our linear methods to more templates than might be required by underlying physical conditions. Details can be found in Appendix B.

We have investigated how many components are needed in a quantitative manner. For the rest of this paper we investigate the how many components are needed qualitatively, by examining some of the simpler NMF component sets, limiting our investigation to $NMF_5$-$NMF_{10}$.

### 3.5.2 Analysis of $NMF_5$ to $NMF_{10}$

Figure 3.3 shows each spectral component for sets $NMF_5 - NMF_{10}$. We have ordered the components so that similar components appear in the same order. We note the ordering of components given by NMF is unimportant.

The NMF sets in Figure 3.3 show that many of the components remain similar, despite an increase in the allowed number of components.

The first component contains a dust continuum which peaks at around 24 $\mu$m and contains emission from the Sulphur line [SIV] at 10.51 $\mu$m, the 12.8, 15.6 and 24.3 $\mu$m Neon lines and Oxygen line [OIV] at 25.89 $\mu$m, all of which are associated with a hot ionised gas source. The continuum in the component from $NMF_9$ and $NMF_{10}$ varies from the others in that continuum does not start until 13 $\mu$m. This coincides with the appearance of the ninth and tenth components which show similar features. The hot dust continuum peaks at a wavelength similar to that of AGN tori, while the hot ionised gas emission lines have also typically been associated with AGN. The appearance of both in one component is consistent with the idea they are correlated.

The second component shows silicate emission features at 10 and 18 $\mu$m due to stretching and bending of the Si-O and O-Si-O bonds respectively. Silicate emission is typically associated with emission from very hot dust, found on the inner surface of AGN tori or narrow line regions (Mason et al., 2009).

The third component captures the 6.2, 7.7, 8.6, 11.3, 12.7, 16.4 and 17.0 $\mu$m PAH features, and a cold dust slope at longer wavelength. There is also emission from Argon line [ArII] at 6.89 $\mu$m and Sulphur line [SIII] at 18.71 $\mu$m. Its shape is

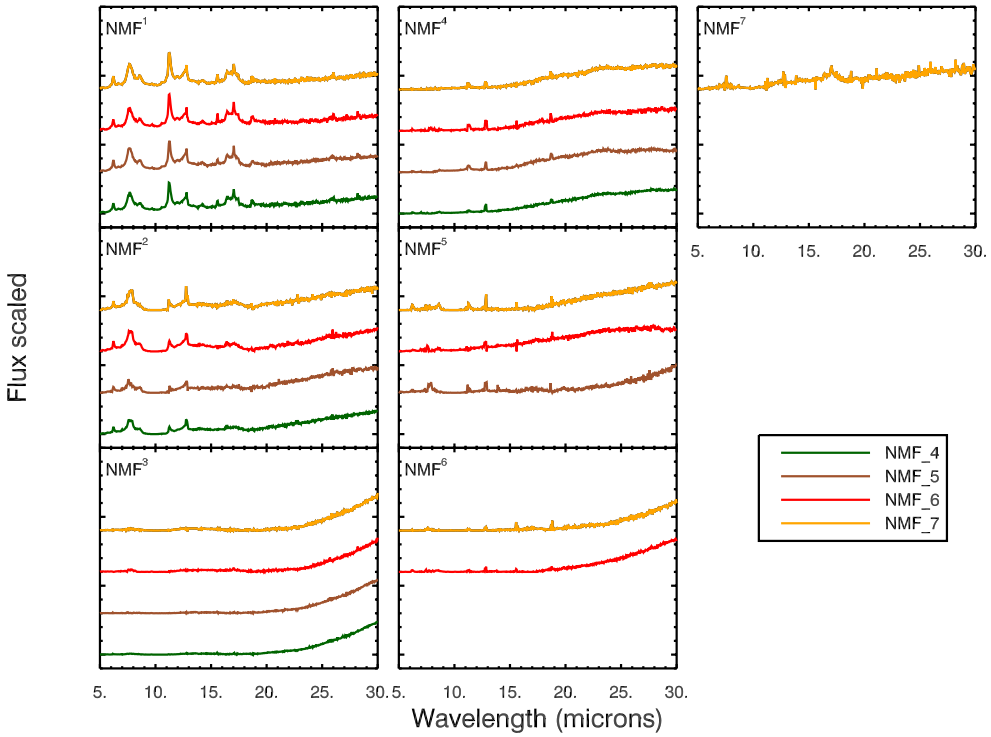


Figure 3.4: The derived NMF spectral components for $NMF_4$-$NMF_7$, using only objects dominated by the third PAH component seen in Figure 3.3. Each NMF set is colour coded, with components ordered by similarity.

similar to the (Brandl et al., 2006) average starburst template, based on 13 starburst galaxies. The ratio of the PAH features are very similar amongst component sets, but dust slope decreases with number of components. The reduction in dust slope for more complex NMF sets coincides with rising continuums seen in the fourth, sixth and seventh components.

The fifth component shows continuum emission up to 7 $\mu$m before dropping off at 10 $\mu$m. It also shows strong emission from the Sulphur line [SIV] at 10.51 $\mu$m. The remainder of the spectrum is noisy and featureless.

The eighth, ninth and tenth components show similarities to the first component. They show varying amounts of emission from the Neon lines, while the merged Oxygen and Iron lines appear as emission in the ninth component and absorption in the tenth. The variation of the first component in $NMF_9$ and $NMF_{10}$ compared to the other NMF sets is a result of the introduction of the ninth and tenth components and occurs because the NMF algorithm is using the freedom of extra components to break down the first into sub components.

**Physical interpretation of the components**

The first two components both show features associated with hot dust and gas emission and are likely to be related to AGN emission. The unified model of AGNs predict silicate emission from type 1 AGN and silicate absorption in type 2 AGN. More recently, the IRS spectra of type 2 quasi-stellar objects (QSOs) have shown silicate emission (Sturm et al., 2006). (Schweitzer et al., 2008) have shown that the IRS spectra of 23 QSOs can be modelled with dusty narrow line region models, while Mason et al. (2009) and Mor et al. (2009) showed that clumpy torus models could also provide silicate emission for both type 1 and type 2 AGN. The fact we see a relatively stable silicate emission component amongst different NMF sets would suggest that silicate emission is occurring in more than just type 1 AGN and is a fundamental spectral component.

The third component is the main star formation component. It is dominated by PAH emission, often used as an indicator of star formation (e.g. Roussel et al., 2001; Peeters et al., 2004; Calzetti et al., 2005; Kennicutt et al., 2009), and predominantly comes from photo-dissociation regions (PDRs) (Roussel et al., 2007; Peeters, 2011). For simpler NMF sets, the component also contains a rising continuum at longer

wavelengths due to colder dust emission (T $\approx 50K$), also associated with star formation (e.g. Calzetti et al., 2007). For the more complex NMF sets, the rising dust continuum is given its own component (e.g. the sixth and seventh). This indicates that although the colder dust and PAH emission both trace star formation, they come from different regions and the NMF algorithm uses the additional freedom of extra components to separate the two. We note that the PAH emission is extremely stable amongst all NMF sets and we do not see significant PAH emission in any other component. Previous studies show the ratio of PAH features vary with metallically and radiation hardness (e.g. Smith et al., 2007), yet we have one component with PAH emission.

To investigate the stability and lack of variation in the PAH emission features, we have re-run the NMF algorithm on objects from our original sample which are dominated by the third component. Figure 3.4 shows the components from $NMF_4$ to $NMF_7$ for our reduced sample. The NMF algorithm now finds two components with PAH emission. The first shows emission at 6.2, 7.7, 8.6, 11.3, 12.7, 16.4 and 17.0 $\mu$m, the second shows reduced emission for the 8.6, 11.3 and 12.7 $\mu$m PAH features and no emission at 16.4 and 17.0 $\mu$m, while at longer wavelengths there is a rising continuum. The two new PAH components show a resemblance to those found in an NMF analysis of IRS spectro-imagery data for galactic PDRs (Berné et al., 2007). Their first component, interpreted as emission from deep within the PDR, showed broad emission at 6.2, 7.8, and 11.4 $\mu$m and a rising continuum. The second component contained emission from the 6.2, 7.6, 8.6, 11.3, 12.7 and 17.4 $\mu$m PAH features, and was shown to be more dominant in regions closer to the star.

By restricting the sample to objects dominated by star formation, the NMF algorithm does not need to use components to separate out hotter dust from AGN, and uses the additional freedom to separate out the PAH emission. The PAH emission in our original third component is therefore capturing the average PAH emission from galaxies.

Components four, six and seven from Figure 3.3, all contain rising continuums, though with varying slopes and is capturing dust emission at different temperatures. The fact we see numerous components with varying slopes suggests that the colder greybody emission of dust varies considerably amongst galaxies. The seventh component also contains a bump at around 8 and 12 $\mu$m. The bumps help build

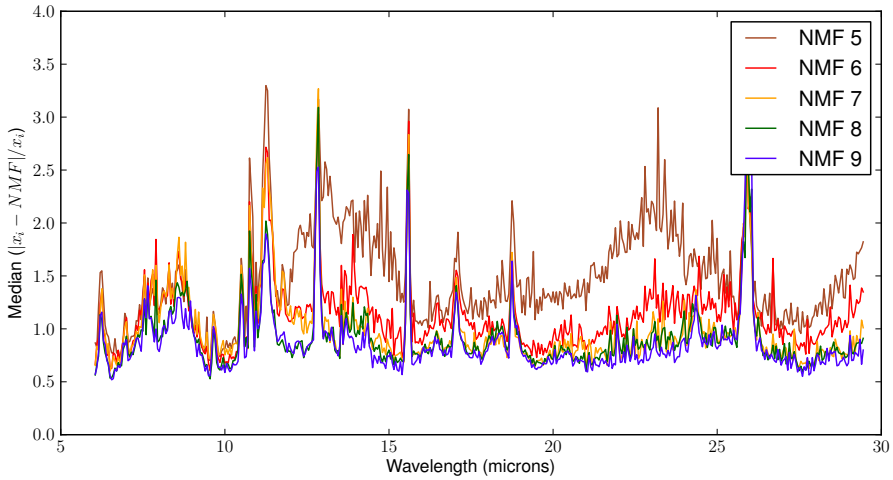


Figure 3.5: The median absolute residuals, normalised by $\sigma$, for NMF sets $NMF_5$-$NMF_9$. The residuals show all NMF sets fail to capture the variance in many of the emission lines. However, for NMF sets $NMF_7$ and above, the residuals for the underlying continuum are down to 1 $\sigma$.

up a silicate absorption feature at 10 $\mu$m, this component is therefore important for dusty galaxies.

To further investigate the components, we can begin to look at how they contribute to different types of spectra. In order to simplify the analysis and to provide a simple set of components, we restrict our components to those from $NMF_7$. Our choice of seven is more qualitative than quantitative, as we have already shown that a quantitative analysis requires more than 14 components. To validate our choice, we have studied the median, absolute residuals of NMF fits to the CASSIS sample with $NMF_5$ to $NMF_9$, shown in Figure 3.5. The residuals are high for some of the emission lines, particularly the PAH features, because our components capture the average line emission. However we note that by seven components, the residuals for the underlying continuum are down to 1 $\sigma$ and there is little advantage in using more complicated sets. By choosing seven, we believe we strike the balance between having enough simplicity to have a useful and physically intuitive NMF set of components, whilst being able to reconstruct the general spectral shape. The seven components are re-plotted in Figure 3.6.

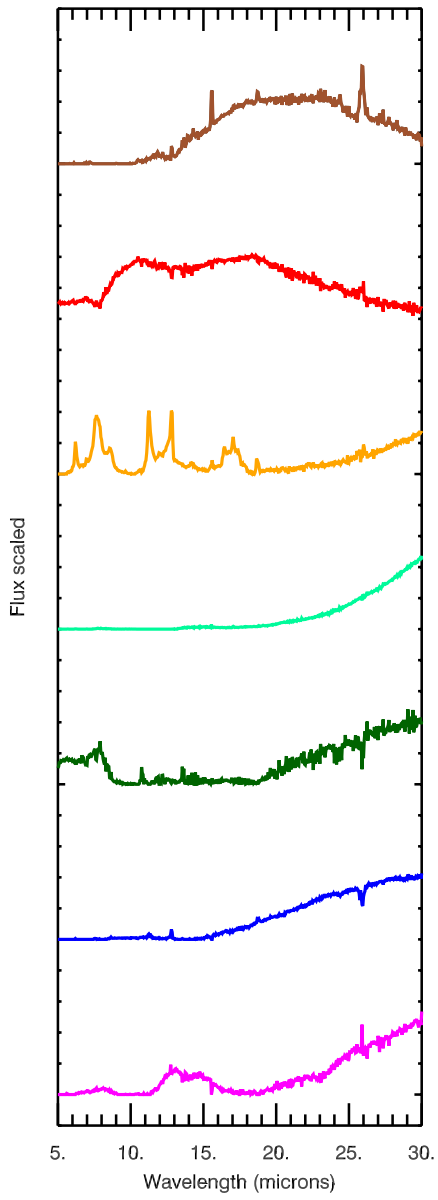


Figure 3.6: The 7 components from $NMF_7$, corresponding to the yellow components in Figure 3.3. The new colour coding is used to identify the different components in subsequent figures.

### $NMF_7$ fits to example galaxy spectra

We now examine the NMF fits to spectra of different types of galaxies in order to show how contributions from components vary and that our $NMF_7$ set can capture the general shape of different types of spectra. Our example fits, along with the corresponding residuals (i.e. data-fit) can be found in Figures 3.17 and 3.18. The first plot in Figure 3.17 shows the NMF fit to the Blue Compact Dwarf (BCD) KUG 1013+381, observed as part of the IRS Guaranteed Time Observation (GTO) program. BCDs tend to be small galaxies with low metallicity, that have undergone a recent burst of star formation but have suppressed star formation compared to typical starburst galaxies (Wu et al., 2006).

Our NMF fit shows component one makes a significant contribution, suggesting there is some hot dust. Component four also makes a large contribution, indicating emission from colder dust. Components six and seven, both containing dust slopes at longer wavelengths, also contribute. There is very little emission from component two, which we believe is associated with the inner surface of an AGN and there is very little emission from the third 'PAH' component. The residual plot shows the $NMF_7$ set can construct the underlying continuum, however the [SIV], [NeIII] and [SIII] emission lines are underestimated.

Our second NMF fit is to the ULIRG and type 1 Seyfert galaxy, Markarian 231. Unlike, KUG 1013+381, the second 'silicate emission' component makes a contribution, and the other, warmer dust components such as six and seven contribute as much power to the longer wavelengths as the fourth component. There is very little contribution from the third component. Residuals show the fit is reasonable except beyond 25 $\mu$m, where there appears to be some instrumental artefact in the spectrum.

The third fit is to PG 1211+143, also a type 1 Seyfert galaxy. The second component dominates the emission of this object. The first, fifth and sixth component make comparable contributions. The residual plot shows that our $NMF_7$ set slightly over estimates emission from the [NeIII] and [OIV] lines.

The fit to the ULIRG and type 2 Seyfert galaxy, Markarian 273, is dominated by emission from the fourth 'cold dust' component. Residuals show the NMF components underestimate some of the emission lines, particularly the [NeIII] line. The continuum appears to be well reconstructed by the NMF components.

Our final two fits in Figure 3.17 are to the starburst galaxies, NGC 3301 and NGC 3256. The third component contributes in the shorter wavelengths, while the colder dust components, four and six, contribute at longer wavelengths. The residuals show the components are capable of reconstructing the continuum, but fail to capture the emission lines accurately.

Four additional example fits are shown in Figure 3.18. The first is to LINER, 3C270. The first, second and fifth components are the main contributors. while the residuals show the fit can reconstruct the continuum, but underestimate the 12.8 $\mu$m Neon line. The submillimeter galaxy GN26 is over a short wavelength region and the spectrum is quite noisy. Our final two fits are to quasar PG0804+761 and ULIRG IRAS 10378+1108. As with other type 1 AGN, the second component dominates emission. Our $NMF_7$ set fails to model the full width of the very broad silicate emission feature at 9.7 $\mu$m, however the rest of the continuum is well reconstructed. Our NMF fit to the ULIRG IRAS 10378+1108 dominates the emission, while the residuals show the NMF set slightly overestimate the greybody emission longwards of 27 $\mu$m.

In addition to galaxy spectra, we also fit our $NMF_7$ set to the average spectral templates from the IRS spectral ATLAS of galaxies (Hernán-Caballero & Hatziminaoglou, 2011). Table 3.3 in Appendix C gives more details on the sources used for the ATLAS average templates. As can be seen in Figure 3.19, the change in contributions for different types of object is consistent with those in Figure 3.17. The continuum is well constructed for all average templates, however the residuals show the emission lines are not accurately reconstructed, especially for the average LINER template.

Overall, our fits show for Seyfert galaxies, the first and second component, along with the warmer dust components of five, six and seven are all important, though their contributions vary. For the starburst galaxies, the third and fourth component play a more important role. The residual plots show that our NMF set is capable of reconstructing the continuum to a reasonable accuracy, however some of the emission lines are not always fitted well. This is to be expected since, as we have previously shown, the components capture the 'average emission' of spectral lines. To accurately fit continuum and emission lines, our Bayesian evidence calculation has shown we would need an NMF set with more than 14 components. The goal of

this paper is to find a physically intuitive component set, which requires a balance between number of components and ability to reconstruct spectra. We believe Figure 3.5 and 3.17 shows our $NMF_7$ set fits this requirement.

To illustrate how the components contribute to a number objects, we can use the weightings provided by the NMF fits as multidimensional co-ordinates. Each galaxy is now a point in a seven dimensional space we call NMF space. We use classifications from the IRS spectral ATLAS of galaxies (Hernán-Caballero & Hatziminaoglou, 2011) to investigate what regions of NMF space are associated with different types of galaxies. The ATLAS collection contains spectra from a number of observing programs. They provide optical classifications from the literature and three additional MIR classifications: MIR SB, MIR AGN1, MIR AGN2 based on the fractional contribution from a PDR component used during spectral decomposition. The AGN subgroups MIR AGN1 and MIR AGN2 are subsets of AGN, classified by whether spectra show silicate emission or silicate absorption. Figure 3.7 shows how objects from the ATLAS groups: MIR AGN1, MIR AGN2, MIR SB, Sbrst, Sy1 and Sy2 are distributed in the seven dimensional NMF space.

As can be seen in Figure 3.7, the Seyfert 1 and MIR AGN1 objects all lie in a region with low contribution from $NMF_7^1$, high contribution from $NMF_7^2$ and very little contribution from $NMF_7^3$. The Seyfert 2 and MIR AGN2 objects are found in a region with a higher contribution in $NMF_7^1$, less or very little contribution from $NMF_7^2$ and very little contribution from $NMF_7^3$. Starburst like objects on the other hand require little contribution from either $NMF_7^1$ or $NMF_7^2$, and a high contribution from $NMF_7^3$.

We note that the components most influential in separating out the different objects are the components one, two and three. Less influential but still significant are the colder dust components $NMF_7^4$ and $NMF_7^6$. They contribute very little to objects classified as AGN, while the contribution for starbursts show a large variation. This fits in with our earlier interpretation that these two components represent obscured star formation components which vary more than the PAH features seen in $NMF_7^3$. The remaining two components are the least significant. There is a slight difference in contribution between AGN 1 objects and the other two classes, while $NMF_7^7$ separates out type 1 and type 2 objects to a certain extent.

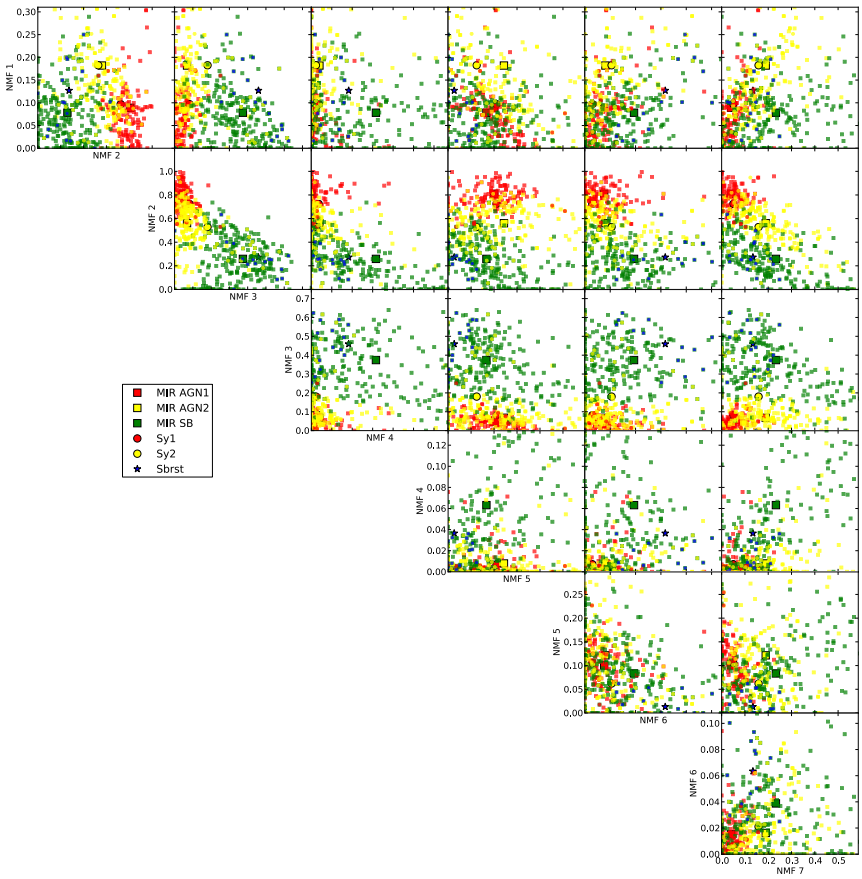


Figure 3.7: The distribution of objects/spectra from the ATLAS groups: MIR AGN1, MIR AGN2, MIR SB, Sbrst, Sy1 and Sy2 in our 7D space defined by the $NMF_7$ set. Symbols and colours for the different groups are described in the legend. The position of the average template for each group is marked by a larger symbol.

### 3.5.3 Gaussian Mixtures Modelling

We have shown NMF space is capable of separating out different types of objects. We now model how objects separate out in this multidimensional space by applying the parametric technique Gaussian mixtures modelling (GMM). GMM has already been successfully applied to the colour and redshift space of galaxies (Davoodi et al., 2006). GMM assumes the distribution of objects can be modelled by a series of clusters, each described by a multidimensional Gaussian. We use the GMM software from the Auton Lab *(Moore, 1999) to model the distribution of the CASSIS sample in our 7 dimensional NMF space. The software uses the Expectation Maximisation algorithm to learn the position and size of the clusters and uses the Akaike Information Criterion (AIC;Akaike (1974)) to select how many are needed to describe the distribution of objects.

We find that 8 clusters are required to adequately model the distribution. Each cluster describes a probability density function (PDF) for any position in NMF space. By using an objects position in NMF space, we can assign it to one of the 8 clusters.†Table 3.1 shows how some of the ATLAS classified sources are distributed across the 8 clusters, with clusters ordered by their normalisation (i.e. how many objects are in that cluster). As can be see in Table 3.1, the majority of objects are contained within the first five clusters. The normalisations associated with the remaining clusters (i.e. how many objects they capture) are also very small. We therefore use the first five clusters to define a classification scheme.

The location in NMF space of the first five clusters can be seen in Figure 3.8. Each cluster is represented by its 1 sigma contour. The CASSIS sample used for training the Gaussian Mixtures modelling are also plotted.

As can be seen in Figure 3.8 and classifications in Table 3.1, cluster one captures nearly all the Seyfert one galaxies, and some Seyfert two galaxies. Cluster two contains a significant number of objects previously classified as starbursts, while cluster three contains a large proportion of the remaining Seyfert two objects. The

*http://www.autonlab.org

†Every position in NMF space has eight PDF values associated with it (one for each cluster). Using the highest probability density provides the optimal (maximum likelihood) classification. However, since the PDFs overlap, this will not provide the best classification for the population statistics. We therefore take the same approach as Davoodi et al. (2006) and randomly assign each galaxy to a cluster, with probability proportional to the PDF values at the galaxies position in NMF space.

position of cluster four indicates this could be an intermediary group between typical Type one and Type two galaxies. The fifth cluster contains just over a fifth of those objects classified as starbursts in the MIR and no optically classified starbursts. Its position in NMF space also suggests it captures those objects which are dusty starbursts.

We conclude that cluster one is related to Seyfert 1 galaxies, cluster two with starbursts, cluster three with Seyfert two galaxies and cluster four for galaxies showing signs of both Seyfert one and Seyfert two (e.g. Type 1.5). The fifth cluster captures those galaxies which are dusty and obscured. The clusters can be used as a classification scheme by taking any IRS galaxy spectrum, fitting with $NMF_7$ set and using the corresponding weights to identify what cluster the object is associated with.

We compare our classification scheme to the Spoon et al. (2007) diagram, which classified ULIRGs via the strength of their 9.7$\mu$m silicate feature and 6.2 $\mu$m equivalent width. Figure 3.9 shows 89 ULIRGs in the Spoon et al. (2007) diagram, colour coded by our our classification. Seyfert one classified galaxies lie on the far left of the bottom horizontal branch, corresponding to a 1A and 1B Spoon classification, Seyfert two classified galaxies span the horizontal branch and 2B Spoon classification. The starburst classified objects are located in the far bottom right of the Spoon diagram, while dusty objects are spread out across the diagonal branch. Only three objects are classified as Type 1.5 and they lie on the horizontal branch, in-between the Seyfert one and Seyfert two classified galaxies.

Comparing the success rates of different classification schemes, without knowing the 'true' classification is always problematic, however our classification scheme is consistent with the Spoon et al. (2007) interpretation of Figure 3.9 in terms of the location of starbursts, AGN dominated objects and dusty objects. Unlike the Spoon diagram, our classification scheme can also distinguish between Seyfert one and Seyfert two galaxies.

We have shown our classification scheme is just as successful as the Spoon classification. However, our classification has three distinct advantages over Spoon et al. (2007). First, Spoon et al. (2007) only use the 9.7$\mu$m silicate feature and 6.2 $\mu$m PAH equivalent width to separate out classes. By using the NMF components as a basis for our GMM based classification scheme, we make use of the whole MIR region to classify objects. This also enables us to classify objects where the 9.7$\mu$m

Table 3.1: The percentage of ATLAS classified objects in each cluster for 7 NMF templates. The first column indicates the cluster number. The second column shows the probability that a CASSIS object is in that cluster (i.e. how many objects can be found in it). The remaining columns contain the percentage of ATLAS classification in each cluster.

| GMM and ATLAS classification for 7 templates | | | | | | | |
|---|---|---|---|---|---|---|---|
| Cluster | prob. | Sy1 | Sy2 | MIR AGN1 | MIR AGN2 | MIR SB | Sbrst |
| 1 | 0.301 | 90.9 | 37.7 | 97.5 | 52.3 | 1.6 | 6.2 |
| 2 | 0.287 | 0.0 | 17.0 | 0.0 | 1.7 | 43.6 | 68.8 |
| 3 | 0.156 | 0.0 | 28.3 | 0.8 | 24.1 | 11.7 | 12.5 |
| 4 | 0.147 | 9.1 | 17.0 | 0.8 | 8.6 | 21.4 | 6.2 |
| 5 | 0.080 | 0.0 | 0.0 | 0.0 | 8.0 | 20.2 | 0.0 |
| 6 | 0.022 | 0.0 | 0.0 | 0.8 | 2.3 | 0.8 | 6.2 |
| 7 | 0.004 | 0.0 | 0.0 | 0.0 | 1.1 | 0.8 | 0.0 |
| 8 | 0.003 | 0.0 | 0.0 | 0.0 | 1.7 | 0.0 | 0.0 |

silicate feature and 6.2 $\mu$m PAH equivalent width are not available or difficult to measure. Secondly, our classification scheme is modelled on the number density of our CASSIS sample in NMF space. Since our sample contains a large variety of objects, any sample biases will have a small affect on the outcome of our classification scheme. The Spoon classes on the other hand, are chosen based on arbitrary cuts in the 9.7$\mu$m silicate feature and 6.2 $\mu$m PAH equivalent width. Thirdly, because our clusters describe a probability density function, we can give an indication of how likely a galaxy could be found in any one of the five clusters. For example, in Table 3.2 we show the probability of being in any of the five clusters for some famous objects.

We make our classification tool publicly available on the arxiv and at `https://github.com/pdh21/NMF_software/`.

### 3.5.4 SF-AGN contribution

We have shown that the NMF components are capable of distinguishing between the objects showing extreme star formation or AGN activity. We now use them to introduce a diagnostic to quantify the contribution from star formation and AGN. Unlike other diagnostics, ours employ the whole MIR spectrum to disentangle the SF versus AGN contributions, and it is not based on specific features for which we

Table 3.2: The approximate probability of being in one of the five clusters in our GMM based classification scheme.

| Object | Cluster1 Sy1 | Cluster2 Sbrst | Cluster3 Sy2 | Cluster4 Sy1.5 | Cluster5 Dusty SB |
|---|---|---|---|---|---|
| Arp220 | 0.00 | 0.23 | 0.41 | 0.02 | 0.34 |
| Mrk231 | 0.32 | 0.00 | 0.34 | 0.32 | 0.02 |
| PG1211+143 | 0.92 | 0.00 | 0.00 | 0.08 | 0.00 |
| IRAS10565+2448 | 0.00 | 0.71 | 0.25 | 0.00 | 0.04 |
| IRAS10378+1109 | 0.00 | 0.01 | 0.06 | 0.00 | 0.93 |

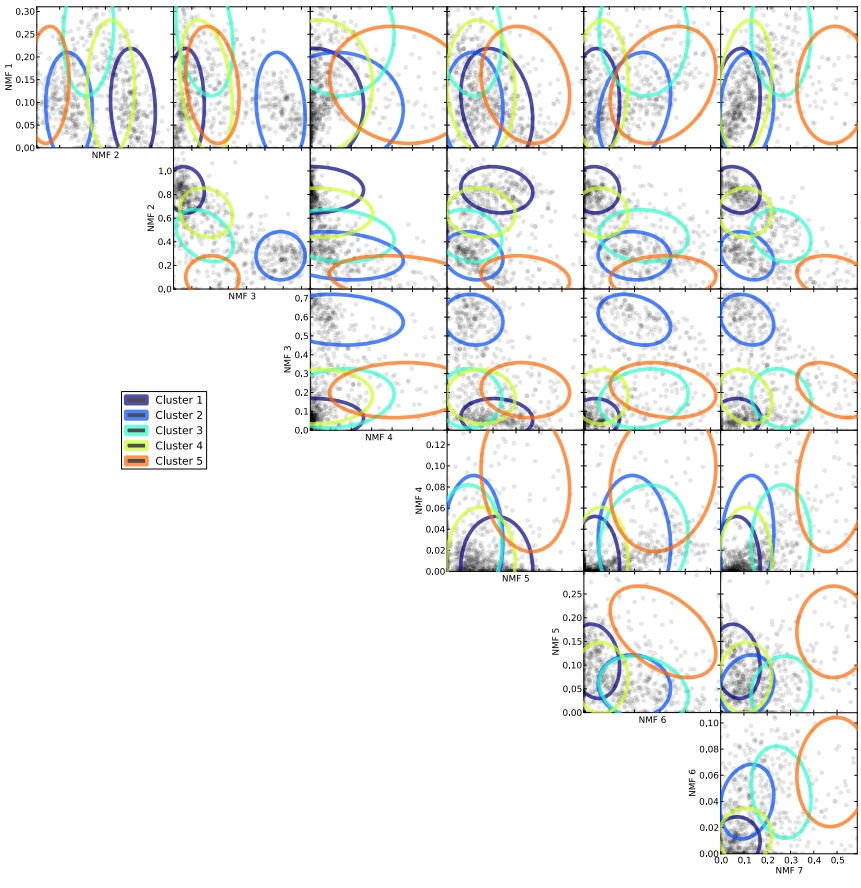


Figure 3.8: NMF space for 7 templates. CASSIS objects used for NMF and GMM are also plotted. The ellipses represent the different clusters found through Gaussian Mixtures Modelling

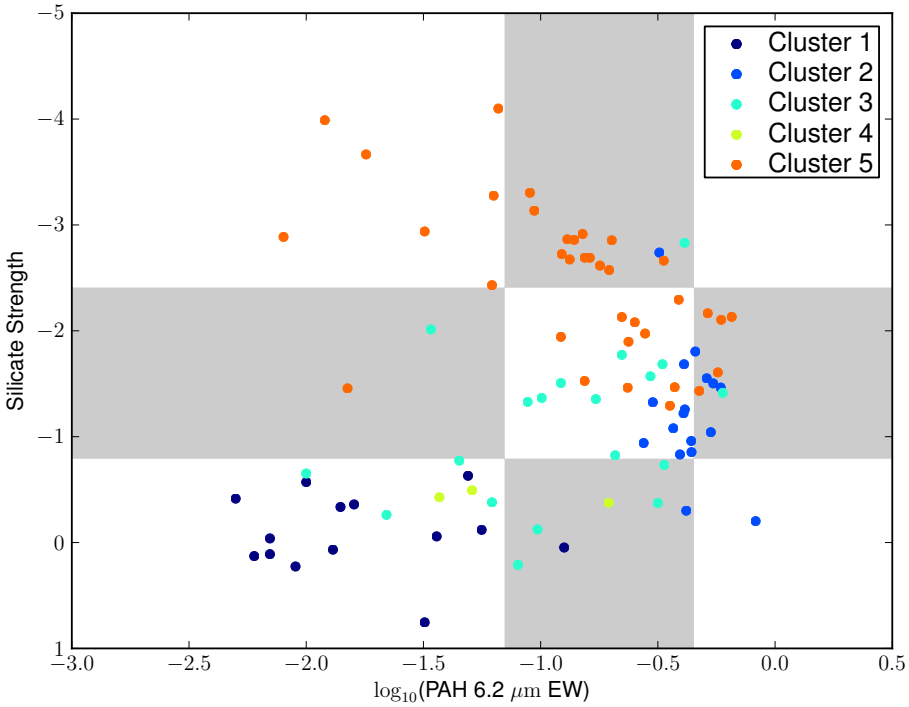


Figure 3.9: The Spoon et al. (2007) diagram showing Silicate strength versus the 6.2 $\mu$m PAH equivalent width. The plot is separated into the different Spoon classes and objects are colour coded by our GMM classification.

need to know information on their origin.

For AGN, $NMF_7^1$ and $NMF_7^2$ are the most important and bear the physical features we know to originate from AGN tori. We therefore adopt $NMF_7^1$ and $NMF_7^2$ as contribution from AGN. For star formation, the third component is the most important, however we argue that the fourth and fifth components are also required as they contain the colder dust associated with obscured star formation. This is especially important for objects like Arp 220 which are known to be predominantly powered by star formation but have less than average PAH emission compared to other submillimeter galaxies (Pope et al., 2008). We do not include $NMF_7^6$ and $NMF_7^7$ in our diagnostic. These components contribute to both AGN and starbursts and we have interpreted them as arbitrary dust components that are not specifically associated with star formation or AGN activity. Our diagnostic is taken as the ratio of MIR luminosity from the following components:

$$\frac{\text{starformation}}{\text{AGN}} = \frac{L_{NMF3} + L_{NMF4} + L_{NMF6}}{L_{NMF1} + L_{NMF2}} \tag{3.10}$$

**Comparison to other MIR diagnostics**

We now show this diagnostic compared to other MIR diagnostic plots quantifying star formation and AGN contribution.

Farrah et al. (2009) applied Bayesian inferencing and graph theory to a data set of 102 mid-infrared spectra. By examining how position in the network was related to other parameters (e.g.infrared luminosity, optical spectral type and black hole mass) they concluded that the network depicted the evolutionary scheme of ULIRGs, with different branches relating to Starburst+AGN and luminous AGN.

We now investigate how our $NMF_7$ set relates to the same network by decomposing the Farrah et al. (2009) sample with our NMF components and colour-coding the network by our NMF diagnostic. The connections are taken from Farrah et al. (2009) and we use the same Cytoscape software * to produce the network. We note that our network is not identical to that in Farrah et al. (2009) due to the random seed starting position used by the spring-embedded algorithm in Cytoscape. The two main branches seen in Farrah et al. (2009) are still seen in Figure 3.10, with the lower and right hand branches corresponding to the Starburst+ AGN and Luminous AGN branches respectively. Each galaxy is colour coded by our new NMF

*Available from http://cytoscape.org/.

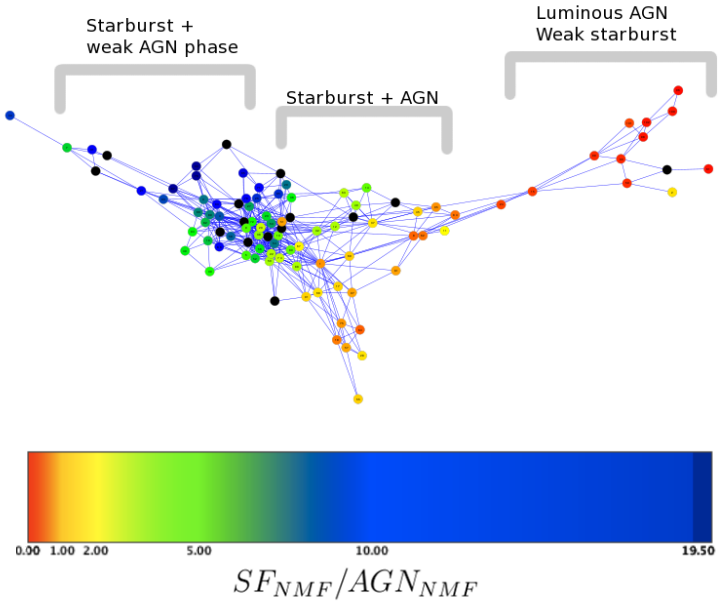


Figure 3.10: The network diagram along with interpretation from Farrah et al. (2009). Starbursts dominate the left hand side of the network. As the AGN becomes more dominant, galaxies move to the right and finally on to one of the two branches. The Nodes are colour coded by our NMF diagnostic. Nodes in black are where spectra are not available.

diagnostic.

As can be seen in Figure 3.10, our NMF diagnostic is consistent with the interpretation that star formation occurs on the left hand side of the network, with AGN activity increasing as we move to the right. The right hand branch appears to be AGN dominated, as was concluded in Farrah et al. (2009).

Our second comparison is with the diagnostic diagram introduced by Laurent et al. (2000) and modified for Spitzer by Armus et al. (2007). The diagrams use the integrated continuum flux from $14-15\mu$m, the integrated continuum flux from $5.3-5.5$ and the $6.2\mu$m PAH flux to indicate fractional contributions from AGN and starbursts. Figure 3.11 shows the same diagnostic plot, plotted with objects from the CASSIS database with measurements of the continuum and $6.2\mu$m flux taken from the CASSIS database. The points are colour coded by our NMF diagnostic.

Objects with a high NMF SF-AGN ratio are located in the top right while objects with a low NMF SF-AGN ratio lie in the bottom left. This is consistent with the simple linear mixing lines indicating AGN and star formation fraction seen in Armus et al. (2007) and Petric et al. (2010).

Our third and fourth comparison is with diagnostic diagrams using emission

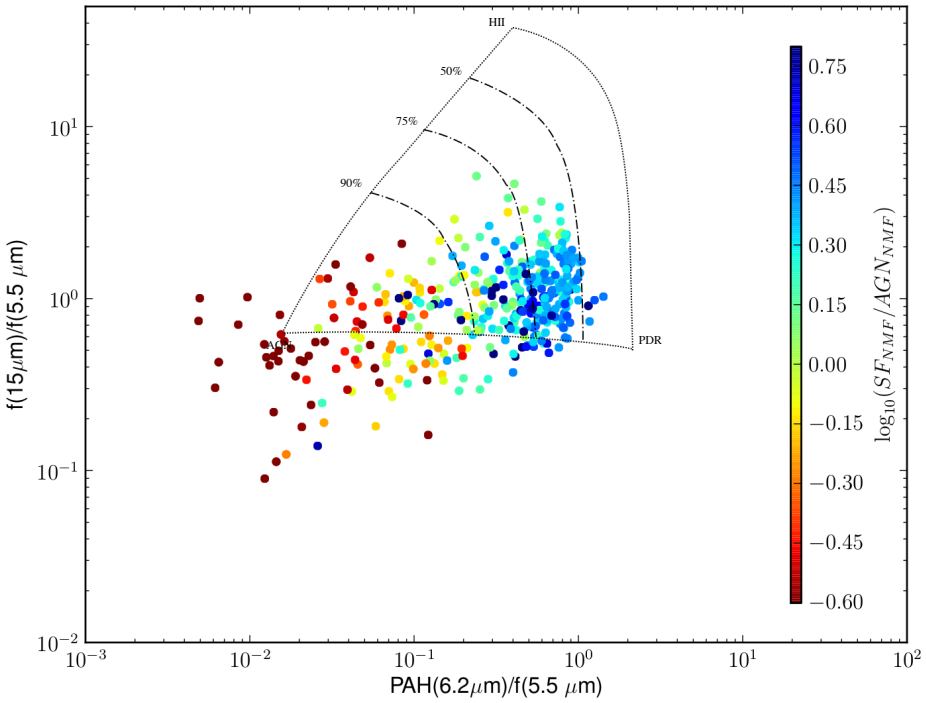


Figure 3.11: The ratio of 15 to 5 $\mu$m continuum flux, against the 6.2$\mu$m PAH flux to 5 $\mu$m continuum flux, as seen in Armus et al. (2007) The three vertices, labeled as AGN, H II, and PDR, represent the positions of 3C 273, M17, and NGC 7023. The dot-dashed lines indicate a 90%, 75%, and 50% fractional AGN contribution, while 100% AGN is defined by the position of 3C 273. Points are colour coded by our NMF diagnostic.

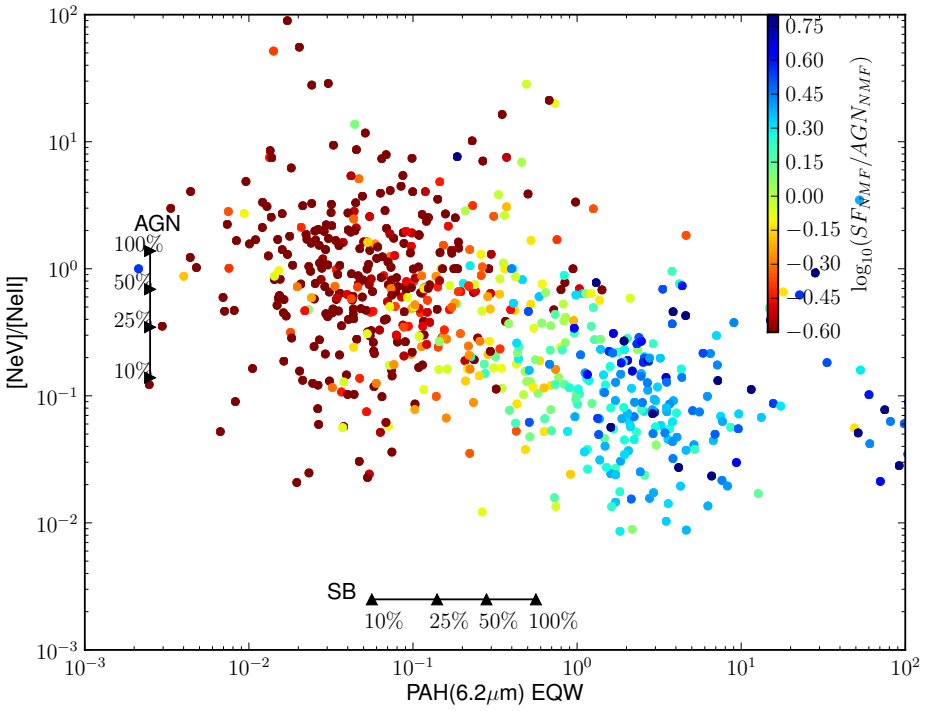


Figure 3.12: The [NeV]/[NeII] ratio vs the PAH 6.2 $\mu$m equivalent width. The points are those objects in the CASSIS database that have a redshift and an estimate for the three lines. The points are colour coded by our NMF diagnostic. We also show the 100%, 50%, 25%, and 10% AGN and starburst linear mixing contributions taken from Armus et al. (2007)

.

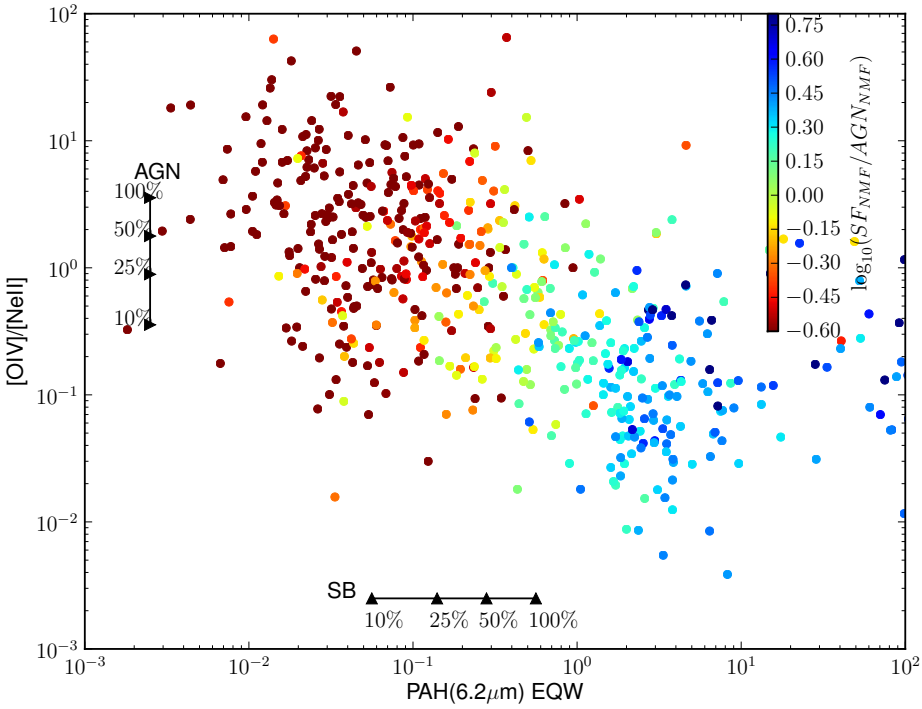


Figure 3.13: The [OIV]/[NeII] ratio vs the PAH 6.2 $\mu$m equivalent width. The points are those objects in the CASSIS database that have a redshift and an estimate for the three lines. The points are colour coded by our NMF diagnostic. We also show the 100%, 50%, 25%, and 10% AGN and starburst linear mixing contributions taken from Armus et al. (2007)

lines. We plot all spectra in the CASSIS database that have a known redshift and measurable emission line. Line measurements are made with the PAHfit software (Smith et al., 2007). Figure 3.12 shows the ratio of Neon forbidden lines [NeV] and [NeII] against the PAH 6.2 $\mu$m equivalent width, colour coded by the NMF diagnostic. We indicate the fractional AGN and starburst contribution to the MIR luminosity from the [NeV]/[NeII] (vertical) and 6.2 $\mu$m PAH EQW (horizontal) assuming a simple linear mixing model. In each case, the 100%, 50%, 25%, and 10% levels are marked. The 100% level is set by the average detected values for the [NeV]/[NeII] and PAH 6.2 $\mu$m equivalent width among AGN and starbursts respectively, as discussed in Armus et al. (2007).

We see that our diagnostic is consistent with star formation dominated objects being located in the bottom right of the plot, while objects with higher AGN contribution are located in the top left.

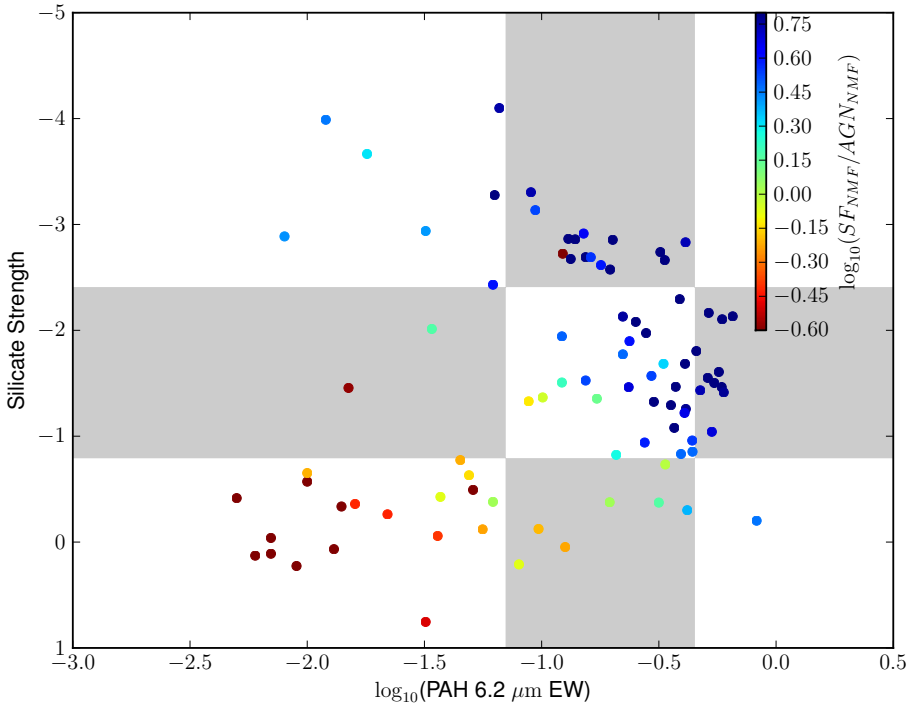


Figure 3.14: The Spoon et al. (2007) diagram showing Silicate strength versus the 6.2 $\mu$m PAH equivalent width. The plot is separated into the different Spoon classes and objects are colour coded by the NMF diagnostic.

The third diagnostic diagram uses the [OIV] and [NeII] ratio vs PAH 6.2 $\mu$m equivalent width. As in Figure 3.12, we colour code the points by NMF diagnostic and indicate the fractional AGN and starburst contributions as discussed in Armus et al. (2007). Our plot can be seen in Figure 3.13. AGN dominated objects lie the top left, star formation dominated objects in the bottom right, which is consistent with the interpretation of Armus et al. (2007). Our final comparison is with Spoon et al. (2007) diagram, classifying ULIRGs via the strength of their 9.7$\mu$m silicate feature and 6.2 $\mu$m equivalent width. Figure 3.14 shows 89 ULIRGs in the Spoon et al. (2007) diagram, colour coded by our NMF diagnostic. Our NMF diagnostic suggests AGN dominated objects are on the horizontal branch, while objects on the diagonal branch appear to have significant activity from star formation and AGN. Objects dominated by star formation lie at the extreme right of the two branches. Our diagnostic is consistent with the interpretation of Spoon et al. (2007).

We have shown that our diagnostic for determining the AGN/star formation ratio is consistent with MIR diagnostic diagrams already in use. Our diagnostic

however has the advantage that it uses a far greater wavelength range than current diagnostics and does not rely on specific line measurements. By using 5 of the 7 components in $NMF_7$, our diagnostic is also flexible enough to account for the difference in spectra amongst star formation or AGN dominated objects.

## 3.6 Conclusions

We have carried out the first empirical attempt at learning the fundamental MIR spectral components of galaxies via the multivariate analysis technique, NMF. We have chosen NMF as the most appropriate matrix factorisation technique for our problem as the non negative constraints required by the algorithm, more closely resembles the physical process of emission in the MIR than techniques used in previous studies (Wang et al., 2011; Hurley et al., 2012). The NMF algorithm has been applied to 729 galaxy spectra, taken from the CASSIS database (Lebouteiller et al., 2011) with spectral redshifts ranging from ($0.01 < z < 0.2$).

We have investigated the number of components needed to accurately reconstruct spectra by evaluating the Bayesian evidence with the nested sampling routine, MULITNEST. The Bayes factor suggests that the number of components exceeds 14 but the gain in increasing the number of components decreases dramatically from seven components onwards. An NMF set with a large number of components may accurately reconstruct all spectra, but assigning physical interpretation to each component becomes difficult, limiting its practical utility.

We have therefore examined the simpler component sets $NMF_5$-$NMF_{10}$. We find that despite an increase in the allowed number of components, many of the components remain similar. For example, similar counterparts to components in $NMF_5$ can be found in $NMF_6$ and above, the sixth component in $NMF_6$ can be found in $NMF_7$ and above and so on. Finding similar components, despite an increase in flexibility, suggests these components are fundamental spectral components.

We find the components also have clear, physical interpretation. The first component contains the forbidden fine structure lines associated with narrow line regions and AGN as well as a hot dust continuum also typical of AGN tori. The second common component shows silicate emission at 10 and 18 $\mu$m and is indicative of the warm dust associated with both the inner wall of the AGN torus or narrow line region clouds. The third component is a star formation component, containing

all of the PAH and molecular hydrogen emission lines, found near PDRs. As the number of components is increased, the colder dust slope is removed to the sixth and seventh components. We interpret this as the separation of unobscured star-forming component (or PDR) from an obscured star-forming component showing colder dust.

Re-running the NMF algorithm on objects dominated by star formation, we show that the PAH emission begins to separate out into two components, which show similar features to the two different PDR components found in Berné et al. (2007).

We have shown that a simpler NMF set with seven components is capable of reproducing the general continuum shape for variety of extragalactic spectra seen in the MIR, though the components struggle with the variation in emission lines. By examining the contributions each component makes to well known objects and previously classified samples, we find different types of objects lie in different regions of 'NMF space'.

Using Gaussian Mixtures modelling, we provide a classification scheme that uses all seven components to separate objects into five different clusters: A Seyfert one cluster, Seyfert two cluster, starburst cluster, dusty and obscured cluster and a type 1.5 Seyfert cluster. Our classification outperforms the Spoon diagram in separating out Seyfert one and two like objects. Unlike the SPoon classification, ours use the whole MIR region, allowing objects without the 9.7$\mu$m silicate feature and 6.2 $\mu$m equivalent width to be classified. Our GMM based classification can also provide an estimate of the probability of finding a particular galaxy in one of the five clusters.

We also use five of the components to create a star formation/AGN diagnostic which performs well against current MIR diagnostic diagrams. Our NMF based diagnostic has the advantage of considering a greater wavelength range, and can therefore be used for objects where specific emission features have not been observed, or for where spectra are too noisy.

Our NMF components provide fundamental, physical components which are ideal for separating out different types of objects and investigating the power associated with AGN and star formation. They are linked to the actual physical environments such as AGN and star formation unlike templates based on specific objects (e.g. M82) or average templates based on a sample of galaxies. We believe our NMF

set could be used to predict useful measures such as star formation rate and AGN luminosity and will investigate this in a future paper. We also believe our NMF set is ideal for more galaxy evolution based investigations such as decomposing the MIR luminosity function into contribution from AGN and and star formation. Our NMF components and code for classification are made available at `https://github.com/pdh21/NMF_software/` and on the arxiv.

## Acknowledgements

We thank the referee for the useful comments, which have improved the paper. We acknowledge support from the Science and Technology Facilities Council [grant numbers ST/F006977/1, ST/I000976/1]. This work is based on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA.

# Bibliography

Akaike H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control*, 19:716. Cited on 83, 103

Alonso-Herrero A., et al. (2011). Torus and agn properties of nearby seyfert galaxies: Results from fitting ir spectral energy distributions and spectroscopy. *arXiv*, astro-ph.CO, 1105.2368v1. Accepted for publication in ApJ. Cited on 60

Alonso-Herrero A., Pereira-Santaella M., Rieke G. H., Rigopoulou D. (2012). Local luminous infrared galaxies. ii. active galactic nucleus activity from spitzer/infrared spectrograph spectra. *The Astrophysical Journal*, 744:2. Cited on 65

Armus L., et al. (2007). Observations of ultraluminous infrared galaxies with the infrared spectrograph on the spitzer space telescope. ii. the iras bright galaxy sample. *The Astrophysical Journal*, 656:148. Cited on 60, 65, 89, 90, 91, 92, 93

Berné O., et al. (2007). Analysis of the emission of very small dust particles from spitzer spectro-imagery data using blind signal separation methods. *Astronomy and Astrophysics*, 469:575. Cited on 62, 76, 95

Blanton M. R., Roweis S. (2007). K-corrections and filter transformations in the ultraviolet, optical, and near infrared. *Astron.J.*, 133:734–754, astro-ph/0606170v1. Cited on 62, 67, 68

Brandl B. R., et al. (2006). The mid-infrared properties of starburst galaxies from spitzer-irs spectroscopy. *The Astrophysical Journal*, 653:1129. Cited on 59, 60, 75

Bromley B. C., Press W. H., Lin H., Kirshner R. P. (1998). Spectral classification and luminosity function of galaxies in the las campanas redshift survey. *The Astrophysical Journal*, 505:25. Cited on 61

Calzetti D., et al. (2005). Star formation in ngc 5194 (m51a): The panchromatic view from galex to spitzer. *The Astrophysical Journal*, 633:871. Cited on 75

Calzetti D., et al. (2007). The calibration of mid-infrared star formation rate indicators. *The Astrophysical Journal*, 666:870. Cited on 76

Chiar J. E., Tielens A. G. G. M. (2006). Pixie dust: The silicate features in the diffuse interstellar medium. *The Astrophysical Journal*, 637:774. Cited on 103

Connolly A. J., Szalay A. S., Bershady M. A., Kinney A. L., Calzetti D. (1995). Spectral classification of galaxies: an orthogonal approach. *Astronomical Journal v.110*, 110:1071. Cited on 61

Dale D. A., et al. (2006). Mid-infrared spectral diagnostics of nuclear and extranuclear regions in nearby galaxies. *The Astrophysical Journal*, 646:161. Cited on 59

Davoodi P., et al. (2006). Parametric modeling of the 3.6-8 m color distributions of galaxies in the swire survey. *The Astronomical Journal*, 132:1818. Cited on 83

Engelbracht C. W., Gordon K. D., Rieke G. H., Werner M. W., Dale D. A., Latter W. B. (2005). Metallicity effects on mid-infrared colors and the 8 m pah emission in galaxies. *The Astrophysical Journal*, 628:L29. Cited on 60

Farrah D., et al. (2007). High-resolution mid-infrared spectroscopy of ultraluminous infrared galaxies. *The Astrophysical Journal*, 667:149. Cited on 60

Farrah D., et al. (2008). The nature of star formation in distant ultraluminous infrared galaxies selected in a remarkably narrow redshift range. *The Astrophysical Journal*, 677:957. Cited on 60

Farrah D., et al. (2009). An evolutionary paradigm for dusty active galaxies at low redshift. *The Astrophysical Journal*, 700:395. Cited on xi, 60, 88, 89

Farrah D., et al. (2012). Direct evidence for termination of obscured star formation by radiatively driven outflows in reddened qsos. *The Astrophysical Journal*, 745:178. Cited on 66

Feroz F., Hobson M. P., Bridges M. (2008). Multinest: an efficient and robust bayesian inference tool for cosmology and particle physics. *arXiv*, astro-ph, 0809.3437v1. Cited on 69

Genzel R., et al. (1998). What powers ultraluminous iras galaxies? *Astrophysical Journal v.498*, 498:579. Cited on 59, 60

Hernán-Caballero A., Hatziminaoglou E. (2011). An atlas of mid-infrared spectra of star-forming and active galaxies. *arXiv*, astro-ph.CO, 1101.4794v1. Cited on ix, 80, 81, 108, 109

Houck J. R., et al. (2004). The infrared spectrograph (irs) on the spitzer space telescope. *The Astrophysical Journal Supplement Series*, 154:18. Cited on 60, 62

Hurley P. D., Oliver S., Farrah D., Wang L., Efstathiou A. (2012). Principal component analysis and radiative transfer modelling of spitzer irs spectra of ultra luminous infrared galaxies. *arXiv*, astro-ph.CO, 1205.6312v1. Cited on 61, 66, 94

Kennicutt R. C., et al. (2009). Dust-corrected star formation rates of galaxies. i. combinations of h and infrared tracers. *The Astrophysical Journal*, 703:1672. Cited on 75

Kessler M. F., et al. (1996). The infrared space observatory (iso) mission. *Astron. Astrophys.*, 315:L27. Cited on 59

Laurent O., Mirabel I. F., Charmandaris V., Gallais P., Madden S. C., Sauvage M., Vigroux L., Cesarsky C. (2000). Mid-infrared diagnostics to distinguish agns from starbursts. *Astronomy and Astrophysics*, 359:887. Cited on 59, 89

Lebouteiller V., Barry D. J., Spoon H. W. W., Bernard-Salas J., Sloan G. C., Houck J. R., Weedman D. W. (2011). Cassis: The cornell atlas of spitzer/infrared spectrograph sources. *The Astrophysical Journal Supplement*, 196:8. Cited on 61, 62, 94

Lee D., Seung S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791. Cited on 62

Lee D., Seung S. (2001). Algorithms for non-negative matrix factorization. In *In NIPS*, pages 556–562. MIT Press. Cited on 61, 62, 67, 68

Lutz D., Spoon H. W. W., Rigopoulou D., Moorwood A. F. M., Genzel R. (1998). The nature and evolution of ultraluminous infrared galaxies:a mid-infrared spectroscopic survey. *The Astrophysical Journal*, 505:L103. Cited on 60

Madden S. C., Galliano F., Jones A. P., Sauvage M. (2006). Ism properties in low-metallicity environments. *Astronomy and Astrophysics*, 446:877. Cited on 60

Mason R. E., Levenson N. A., Shi Y., Packham C., Gorjian V., Cleary K., Rhee J., Werner M. (2009). The origin of the silicate emission features in the seyfert 2 galaxy ngc 2110. *The Astrophysical Journal Letters*, 693:L136. Cited on 73, 75

Moore A. (1999). Very fast em-based mixture model clustering using multiresolution kd-trees. In Kearns M., Cohn D., editors, *Advances in Neural Information Processing Systems*, pages 543–549, 340 Pine Street, 6th Fl., San Francisco, CA 94104. Morgan Kaufman. Cited on 83

Mor R., Netzer H., Elitzur M. (2009). Dusty structure around type-i active galactic nuclei: Clumpy torus narrow-line region and near-nucleus hot dust. *The Astrophysical Journal*, 705:298. Cited on 75

Pan B., Lai J., Chen W.-S. (2011). Nonlinear nonnegative matrix factorization based on mercer kernel construction. *Pattern Recognition*, 44:2800 – 2810. Semi-Supervised Learning for Visual Content Analysis and Understanding¡/ce:title¿. Cited on 67

Peeters E., Spoon H. W. W., Tielens A. G. G. M. (2004). Polycyclic aromatic hydrocarbons as a tracer of star formation? *The Astrophysical Journal*, 613:986. Cited on 75

Peeters E. (2011). The pah hypothesis after 25 years. *arXiv*, astro-ph.GA, 1111.3680v1. Cited on 75

Petric A. O., et al. (2010). Mid-infrared spectral diagnostics of luminous infrared galaxies. *arXiv*, astro-ph.GA, 1012.1891v1. Cited on 60, 65, 89

Pope A., et al. (2008). Mid-infrared spectral diagnosis of submillimeter galaxies. *The Astrophysical Journal*, 675:1171. Cited on 60, 65, 88

Rigopoulou D., Spoon H. W. W., Genzel R., Lutz D., Moorwood A. F. M., Tran Q. D. (1999). A large mid-infrared spectroscopic and near-infrared imaging survey of ultraluminous infrared galaxies: Their nature and evolution. *The Astronomical Journal*, 118:2625. Cited on 60

Rosenberg M. J. F., Berné O., Boersma C., Allamandola L. J., Tielens A. G. G. M. (2011). Coupled blind signal separation and spectroscopic database fitting of the mid infrared pah features. *arXiv*, astro-ph.GA, 1106.5899v1. Cited on 62

Roussel H., Sauvage M., Vigroux L., Bosma A. (2001). The relationship between star formation rates and mid-infrared emission in galactic disks. *Astronomy and Astrophysics*, 372:427. Cited on 75

Roussel H., et al. (2007). Warm molecular hydrogen in the spitzer sings galaxy sample. *arXiv*, astro-ph, 0707.0395v1. Cited on 75

Rovilos E., et al. (2012). Goods-herschel: ultra-deep xmm-newton observations reveal agn/star-formation connection. *Astronomy & Astrophysics*, 546:58. Cited on 66

Sajina A., Yan L., Armus L., Choi P., Fadda D., Helou G., Spoon H. (2007). Spitzer mid-infrared spectroscopy of infrared luminous galaxies at z 2. ii. diagnostics. *The Astrophysical Journal*, 664:713. Cited on 65

Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z. (1988). Ultraluminous infrared galaxies and the origin of quasars. *Astrophysical Journal*, 325:74. Cited on 66

Schwarz G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464. Cited on 103

Schweitzer M., et al. (2008). Extended silicate dust emission in palomar-green qsos. *The Astrophysical Journal*, 679:101. Cited on 75

Skilling J. (2004). Nested sampling. *BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. AIP Conference Proceedings*, 735:395. Cited on 69

Smith J. D. T., et al. (2007). The mid-infrared spectrum of star-forming galaxies: Global properties of polycyclic aromatic hydrocarbon emission. *The Astrophysical Journal*, 656:770. Cited on 76, 92

Spoon H. W. W., Marshall J. A., Houck J. R., Elitzur M., Hao L., Armus L., Brandl B. R., Charmandaris V. (2007). Mid-infrared galaxy classification based on silicate

obscuration and pah equivalent width. *The Astrophysical Journal*, 654:L49. Cited on 60, 65, 84, 87, 93

Sturm E., Lutz D., Tran D., Feuchtgruber H., Genzel R., Kunze D., Moorwood A. F. M., Thornley M. D. (2000). Iso-sws spectra of galaxies: Continuum and features. *Astronomy and Astrophysics*, 358:481. Cited on 59

Sturm E., et al. (2006). Mid-infrared diagnostics of liners. *The Astrophysical Journal*, 653:L13. Cited on 75

Taghizadeh-Popp M., Heinis S., Szalay A. S. (2012). Single parameter galaxy classification: The principal curve through the multi-dimensional space of galaxy properties. *The Astrophysical Journal*, 755:143. Cited on 61

Thornley M. D., Schreiber N. M. F., Lutz D., Genzel R., Spoon H. W. W., Kunze D., Sternberg A. (2000). Massive star formation and evolution in starburst galaxies: Mid-infrared spectroscopy with the iso short wavelength spectrometer. *The Astrophysical Journal*, 539:641. Cited on 60

Valiante E., Lutz D., Sturm E., Genzel R., Chapin E. L. (2009). A backward evolution model for infrared surveys: The role of agn- and color-l tir distributions. *The Astrophysical Journal*, 701:1814. Cited on 60

Veilleux S., et al. (2009). Spitzer quasar and ulirg evolution study (quest). iv. comparison of 1 jy ultraluminous infrared galaxies with palomar-green quasars. *The Astrophysical Journal Supplement*, 182:628. Cited on 60

Wang L., Farrah D., Connolly B., Connolly N., LeBouteiller V., Oliver S., Spoon H. (2011). Principal component analysis of the spitzer irs spectra of ultraluminous infrared galaxies. *Monthly Notices of the Royal Astronomical Society*, 411:1809. Cited on 61, 66, 94

Werner M. W., et al. (2004). The spitzer space telescope mission. *The Astrophysical Journal Supplement Series*, 154:1. Cited on 60, 62

Wu Y., Charmandaris V., Hao L., Brandl B. R., Bernard-Salas J., Spoon H. W. W., Houck J. R. (2006). Mid-infrared properties of low-metallicity blue compact dwarf galaxies from the spitzer infrared spectrograph. *The Astrophysical Journal*, 639:157. Cited on 60, 79

Zafeiriou S., Petrou M. (2010). Nonlinear non-negative component analysis algorithms. *IEEE Transactions on Image Processing*, 19:1050. Cited on 67

## 3.7 (Appendix A) Non-linear matrix factorisation techniques

ICA, PCA and NMF are linear models and cannot efficiently model non-linearities such as dust extinction. Over the last decade, non-linear matrix factorisation techniques have been developed to overcome certain non-linear situations. All of these nonlinear based techniques use kernels to map data with nonlinear structure into a kernel feature space, where the structure becomes linear. Techniques such as PCA or NMF can then be performed in the kernel feature space to recover the structure. These types of techniques are suited to problems where the non-linearity is of parametric form, e.g. points distributed along a circle. Dust extinction is exponential relationship unsuited to this type of technique (Binbin Pan, private communication).

## 3.8 (Appendix B) $NMF_{30}$ and Extinction simulation

To explore whether the extinction can cause problems with our NMF analysis, we simulate extinction via equation 3.3 described in section 3.4.

Our simulation is divided into two parts. The first part assumes galaxy spectra are a linear combination as described in equation 3.2, while the second assumes equation 3.3 is valid. To simulate the spectra, we use NMF set $NMF_5$ and linearly combine them with weights randomly sampled from a distribution based on those found in the real sample. We do this 500 times to create 500 unique galaxy spectra.

The second part of our simulation involves adding extinction to the simulated spectra as described in equation 3.3 in section 3.4. $\tau(\lambda)$ is defined by the Galactic Centre extinction law of Chiar & Tielens (2006).

We then carry out the NMF algorithm on both the unextincted and extincted spectra. We run the algorithm for $NMF_5$-$NMF_{20}$ and use the simplified model selection measures: the Akaike Information Criterion (AIC;Akaike (1974)) and the Bayesian Information Criterion (BIC; Schwarz (1978)), defined as follows:

$$AIC \equiv -2 \ln L_{max} + 2k + \frac{2k(k+1)}{N-k-1} \tag{3.11}$$

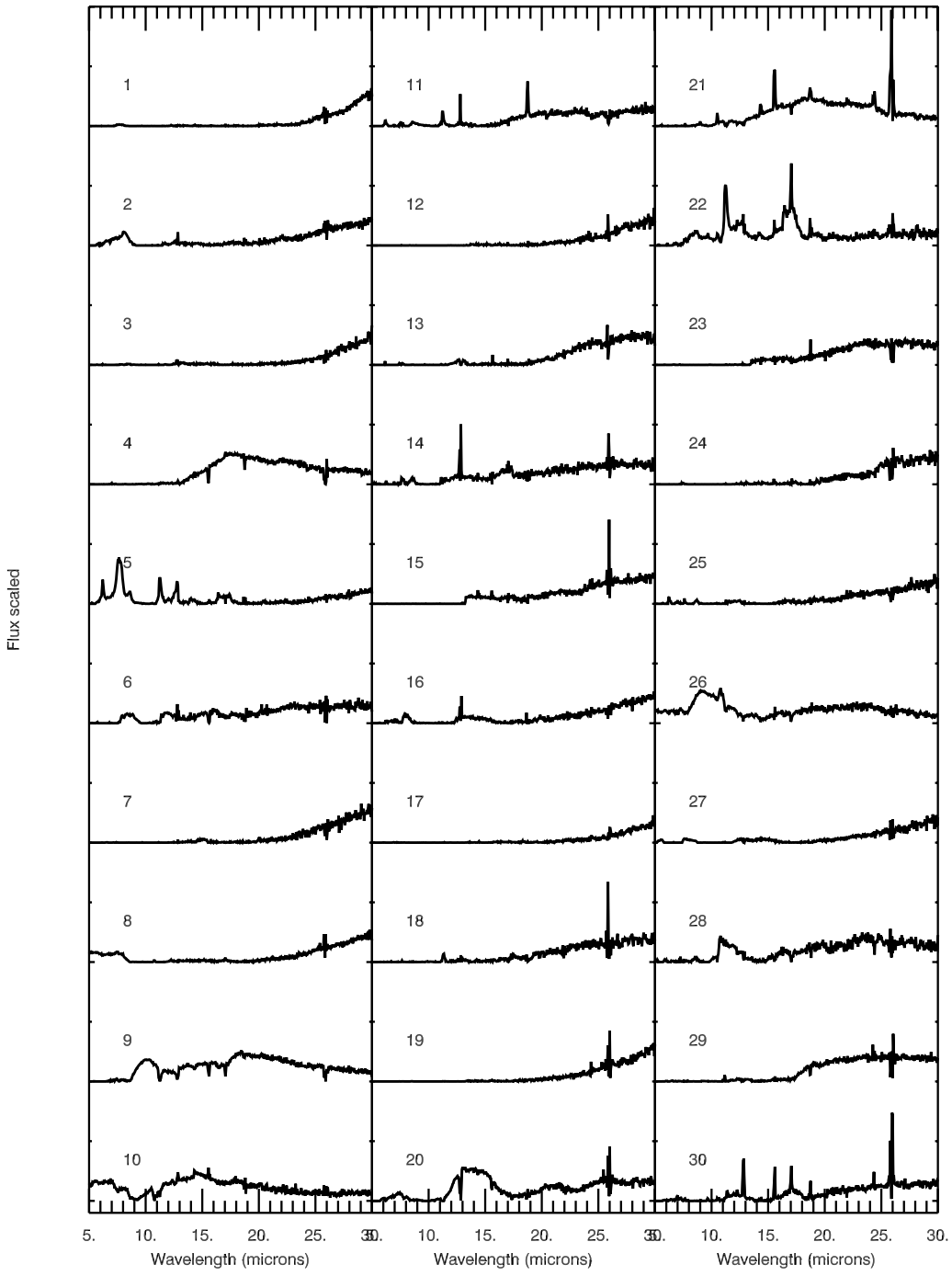


Figure 3.15: The 30 components of NMF set $NMF_{30}$.

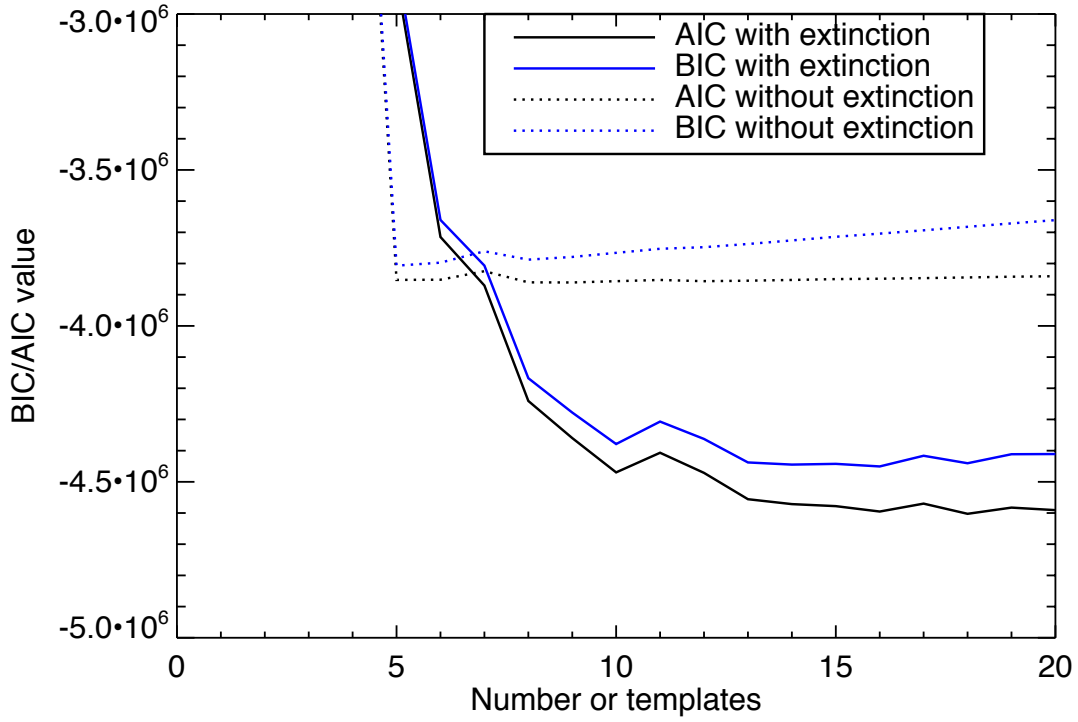


Figure 3.16: The AIC and BIC for the non linear simulations. Both the BIC and AIC for spectra without extinction indicate 5 components as expected. The set with extinction requires around 15-20.

$$BIC \equiv -2 \ln L_{max} + k \ln N \tag{3.12}$$

$L_{max}$ is the maximum likelihood solution, $N$ is the number of datapoints and $k$ is the number of parameters. A minimum value for the AIC and BIC correspond to the optimum model. Figure 3.16 shows both the BIC and AIC for both sets of simulated spectra. As expected, the BIC and AIC indicate the spectra without extinction can be adequately described by the NMF set with 5 components. For spectra with extinction, the BIC and AIC do not level of until $NMF_{15}$-$NMF_{20}$. This suggests that extinction could be a factor in driving our linear methods to more templates than might be required by underlying physical conditions.

## 3.9 (Appendix C) $NMF_7$ fits to galaxy spectra

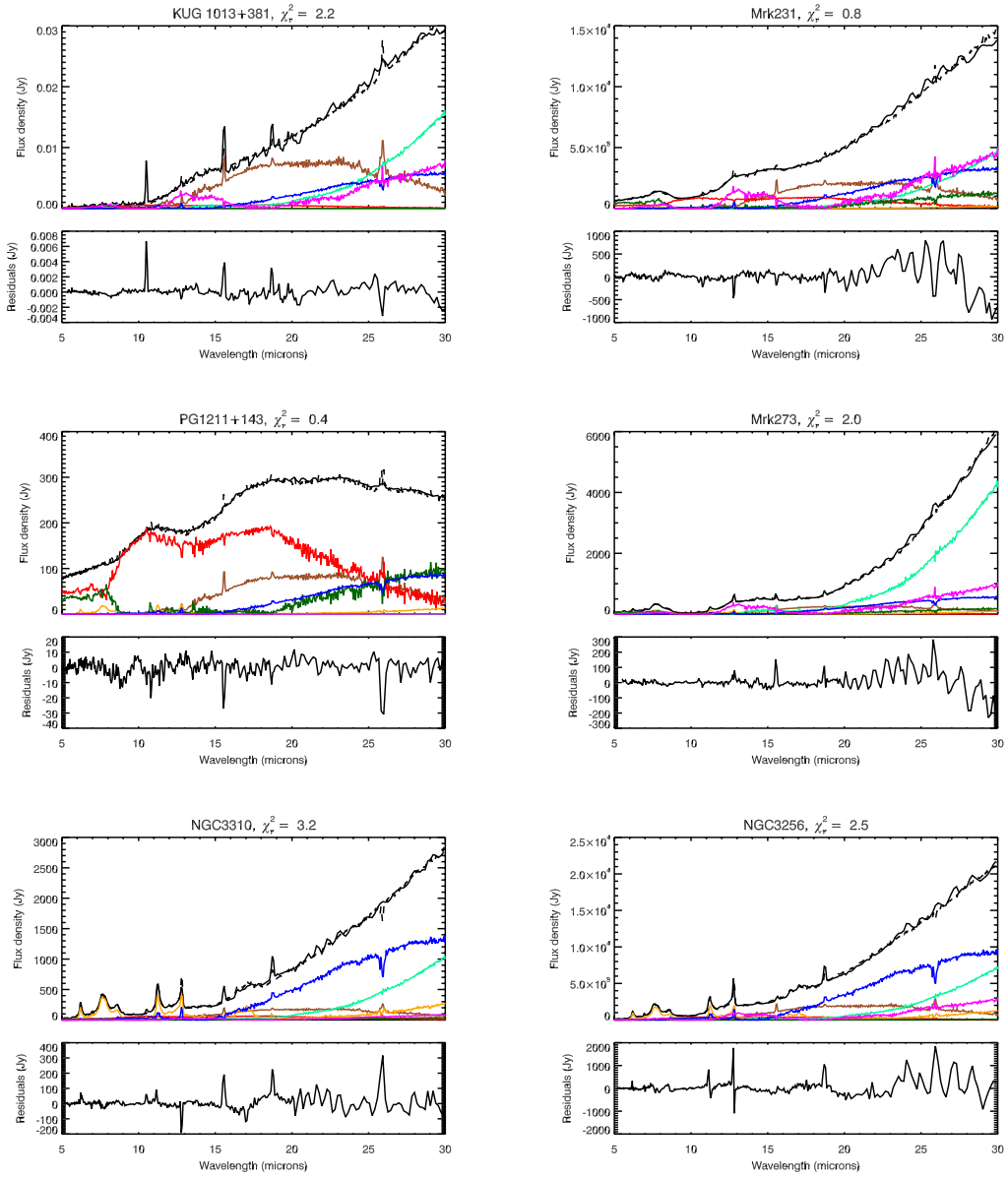


Figure 3.17: $NMF_7$ fits to the Blue Compact Dwarf: KUG 1013+381, Seyfert type 1 galaxies: Markarian 231 and PG1211+143, Seyfert Type 2 galaxy: Markarian 273, and starburst galaxies: NGC3310 and NGC3256. Each spectrum is plotted as a black solid line and the NMF fit as black dashed line. The contribution from each component is also shown, with the same colour coding as in Figure 3.6. The residuals (data-fit) are plotted below each fit.

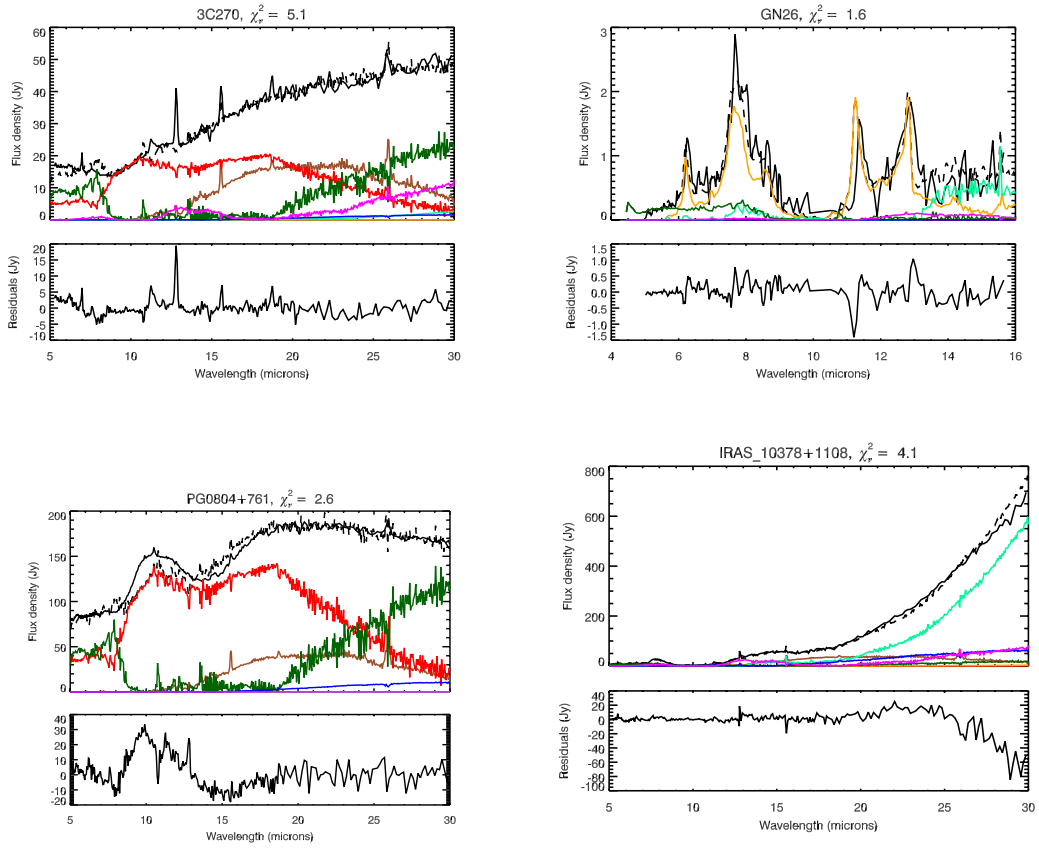


Figure 3.18: Four additional example $NMF_7$ fits. LINER: 3C270, submillimeter galaxy: SMG GN26, quasar: PG0804+761, and ULIRG:IRAS 10378+1108 . Each spectrum is plotted as a black solid line and the NMF fit as black dashed line. The contribution from each component is also shown, with the same colour coding as in Figure 3.6

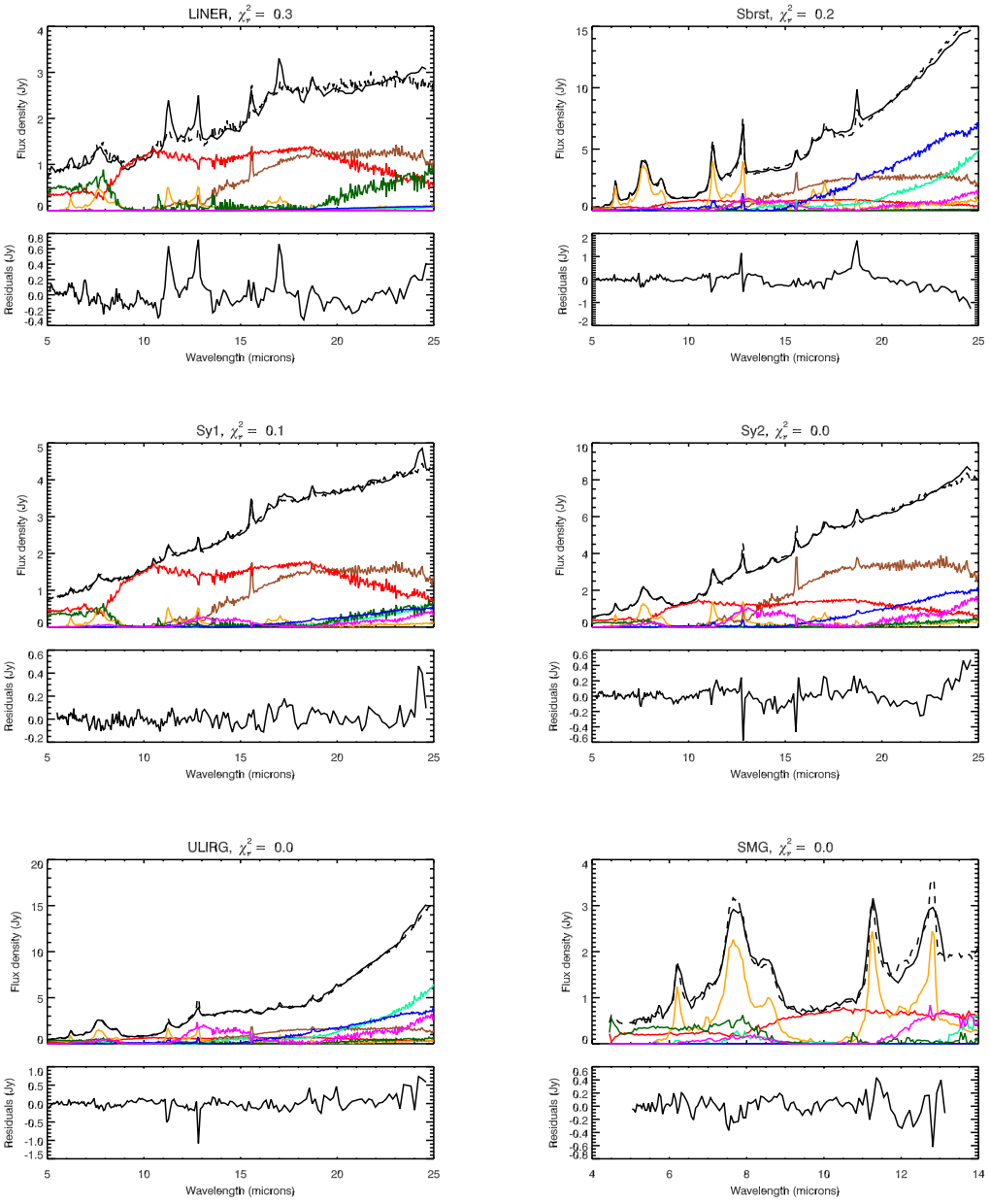


Figure 3.19: $NMF_7$ fits to the Average templates from Hernán-Caballero & Hatziminaoglou (2011) (Information on sample can be found in Table 3.3). Each spectrum is plotted as a black solid line and the NMF fit as black dashed line. The contribution from each component is also shown, with the same colour coding as in Figure 3.6.

Table 3.3: Classification of sources by Hernán-Caballero & Hatziminaoglou (2011)

| name | Nsources | $z_{min}$ | $\langle z \rangle$ | $z_{max}$ | $\lambda_{min}$ [µm] | $\lambda_{max}$ [µm] | comments |
|---|---|---|---|---|---|---|---|
| Sy1 | 11 | 0.002 | 0.041 | 0.205 | 5.2 | 24.6 | Seyfert 1 with $\nu L\nu(7\mu m) < 10^{44}$ erg s$^{-1}$ |
| Sy1x | 72 | 0.003 | 0.091 | 0.371 | 5.0 | 24.6 | intermediate Seyfert types (1.2, 1.5, 1.8, 1.9) |
| Sy2 | 53 | 0.003 | 0.045 | 1.140 | 5.2 | 24.6 | Seyfert 2 with $\nu L\nu(7\mu m) < 10^{44}$ erg s$^{-1}$ |
| LINER | 16 | 0.001 | 0.034 | 0.322 | 5.2 | 24.6 | LINER with $\nu L\nu(7\mu m) < 10^{44}$ erg s$^{-1}$ |
| QSO | 125 | 0.020 | 1.092 | 3.355 | 2.5 | 24.6 | QSO1 and Seyfert 1 with $\nu L\nu(7\mu m) > 10^{44}$ erg s$^{-1}$ |
| QSO2 | 65 | 0.031 | 1.062 | 3.700 | 3.6 | 24.6 | QSO2 and Seyfert 2 with $\nu L\nu(7\mu m) > 10^{44}$ erg s$^{-1}$ |
| Sbrst | 16 | 0.001 | 0.091 | 1.316 | 5.2 | 24.6 | Starburst or HII with $\nu L\nu(7\mu m) < 10^{44}$ erg s$^{-1}$ |
| ULIRG | 184 | 0.018 | 0.730 | 2.704 | 4.5 | 24.6 | ULIRG (low and high redshift sources) |
| SMG | 51 | 0.557 | 1.869 | 3.350 | 4.8 | 12.0 | Submillimiter Galaxies |
| MIR_AGN1 | 119 | 0.002 | 0.455 | 2.190 | 4.0 | 24.6 | MIR selected AGN with silicate emission |
| MIR_AGN2 | 160 | 0.002 | 0.549 | 2.470 | 4.5 | 24.6 | MIR selected AGN with silicate absorption |
| MIR_SB | 257 | 0.001 | 0.413 | 2.000 | 4.6 | 24.6 | MIR selected starbursts |

# Chapter 4

# Herschel-SPIRE-Fourier Transform Spectroscopy of the nearby spiral galaxy IC342

*D. Rigopoulou, P.D. Hurley, B.M.Swinyard, J. Virdee, K.V.Croxall, R.H.B.Hopwood, T.Lim, G.E.Magdis, C.P.Pearson, E. Pellegrini, E.Polehampton, J-D. Smith*

My primary contribution to this paper was the radiative transfer modelling of the CO SLED, described in Sections 4.4 and 4.5. I carried out all modelling of the CO SLED with my combined setup of *Multinest* (Feroz & Hobson, 2008) and RADEX (van der Tak et al., 2007). The associated figures in the two sections were also created by me. I also made significant contributions to both the interpretation and text in these sections. My other contributions include suggestions and corrections to the remaining text throughout the paper.

## 4.1 Abstract

We present observations of the nearby spiral galaxy IC342 with the Herschel Spectral and Photometric Imaging Receiver (SPIRE) Fourier Transform Spectrometer. The spectral range afforded by SPIRE, 196-671 $\mu$m, allows us to access a number of $^{12}$CO lines from J=4–3 to J=13–12 with the highest J transitions observed for the first time. In addition we present measurements of $^{13}$CO, [CI] and [NII]. We use a radiative transfer code coupled with Bayesian likelihood analysis to model and constrain the temperature, density and column density of the gas. We find two

$^{12}CO$ components, one at 35 K and one at 400 K with CO column densities of $6.3\times10^{17}$ cm$^{-2}$ and $0.4\times10^{17}$ cm$^{-2}$ and CO gas masses of $1.26\times10^{7}$ M$_\odot$ and $0.15\times10^{7}$ M$_\odot$, for the cold and warm components, respectively. The inclusion of the high-J $^{12}$CO line observations, indicate the existence of a much warmer gas component ($\sim$400 K) confirming earlier findings from $H_2$ rotational line analysis from ISO and Spitzer. The mass of the warm gas is 10% of the cold gas, but it likely dominates the CO luminosity. In addition, we detect strong emission from [NII] 205 $\mu$m and the $^3P_1 \rightarrow ^3 P_0$ and $^3P_2 \rightarrow ^3 P_1$ [CI] lines at 370 and 608 $\mu$m, respectively. The measured $^{12}$CO line ratios can be explained by Photon-dominated region (PDR) models although additional heating by e.g. cosmic rays cannot be excluded. The measured [CI] line ratio together with the derived [C] column density of $2.1\times10^{17}$ cm$^{-2}$ and the fact that [CI] is weaker than CO emission in IC342 suggests that [CI] likely arises in a thin layer on the outside of the CO emitting molecular clouds consistent with PDRs playing an important role.

## 4.2 Introduction

Far-infrared fine-structure lines of abundant elements such as carbon, oxygen, nitrogen and sulphur either in their neutral or ionised state contribute significantly to the gas cooling of the interstellar medium (ISM, (e.g. Hollenbach & Tielens, 1999)). Far-infrared lines of ionised atoms are useful probes of HII regions while the main cooling of the neutral ISM is carried out by [CII] and [OI] (e.g. Malhotra et al., 2001). In molecular gas, cooling is due to [CI] and the carbon monoxide molecule CO. As potential tracers of the gas cooling, submillimeter [CI] and CO lines are expected to provide information on the gas heating rate, which is dominated by the incident FUV radiation, mainly due to massive and young stars.

Neutral atomic carbon can be found in all types of neutral clouds from diffuse to molecular. The ratio of the two ground state fine-structure lines is a sensitive tracer of the total gas pressure (e.g. Gerin & Phillips, 2000)). Emission from the two ground state fine-structure lines of atomic carbon is seen by COBE throughout the Milky Way and makes a significant contribution to the gas cooling. Despite the high abundance of atomic carbon [CI] in cool interstellar media and the importance in controlling the overall thermal budget only a handful of measurements of the ground state fine structure lines at 370 and 608 $\mu$m have been achieved from the

ground. The first detection was reported by Buettgenbach et al. (1992) in IC342. A handful of galaxies have been detected since, including NGC 253 (Harrison et al., 1995), M82 (Stutzki et al., 1997), M83 (Petitpas & Wilson, 1998), and M33 (Wilson, 1997).

The molecular CO transitions have been extensively studied from the ground. However, the diagnostic power of the CO rotational transitions has not been fully exploited since only the lowest transitions are easily accessible with ground-based telescopes. In recent years higher rotational transitions have been observed in a handful of nearby (mostly) starburst galaxies (e.g. Papadopoulos et al., 2010). The so called "CO Spectral Line Energy Distribution (SLED)" is used to probe the physical properties of the molecular gas such as temperature and column density.

IC342 is a nearby (D=1.8 Mpc, 1"=8.7pc, McCall 1989) spiral galaxy. **Because of its proximity, face-on grand spiral appearance, enhanced star-forming activity in the central region (e.g. Becklin et al., 1980) and strong millimetre and submillimetre emission IC342 has been a popular target for infrared and submillimetre observations. With a far-infrared luminosity $1.25\times10^{10}L_{\odot}$ (Dale et al., 2012), IC342 has been an early target of many investigations of molecular gas and/or atomic far-infrared fine structure lines.** [CI] emission has already been detected from the ground (Buettgenbach et al., 1992) while a number of CO transitions have also been observed (e.g. Bayet et al., 2004, 2006) allowing some constraints to be placed on the properties of the neutral and molecular ISM.

In this paper we present new spectroscopic observations of IC342 obtained using the SPIRE instrument on the Herschel Space Observatory (Pilbratt et al., 2010) covering the 194 to 671 $\mu$m regime. This spectral region is particularly important as it allow us to access a number of high-J CO line transitions and consequently investigate the properties of the molecular gas. In particular, the current observations allow us to probe the peak of the CO SLED and enable modeling of the physical properties of the molecular ISM. In addition the detection of the two ground state [CI] and [NII] lines allow us to investigate the conditions in the PDRs. The current observations demonstrate the power of far-infrared and submillimetre spectroscopy to probe the diffuse and ionised media in external galaxies. The imminent availability of the Atacama Large Millimetre Array (ALMA) will enable this kind of

science in distant galaxies. With a moderate FIR luminosity IC342 is representative of a typical high-redshift galaxy and can serve as a template when designing ALMA observations.

## 4.3 Observations, Data Reduction and Results

The present observations were taken as part of Herschel's Performance Verification (PV) phase using SPIRE (Griffin et al., 2010) as an imaging Fourier-Transform Spectrometer (FTS). The SPIRE astronomical calibration methods and accuracy have been presented in Swinyard et al. (2010). SPIRE-FTS was used in the high spectral resolution mode, sampling across a field of view of 2.6' in diameter. The pointing of the observation was centered at R.A.=56.70322 deg. and Dec=68.09614 deg and the total on-source integration time was 9240 seconds. The SPIRE-FTS measures the Fourier transform of the spectrum of a source using two detector arrays: SSW covering the 194-313 $\mu$m and SLW covering 303–671 $\mu$m wavelength bands simultaneously. The FWHM beamwidths of the SSW and SLW arrays vary from 17" at 194 $\mu$m to 42" at 671 $\mu$m, respectively. The size of the beams varies within this range in a complex fashion due to the nature of the SPIRE detectors (Swinyard et al., 2010; Wu et al., 2013).

The FTS observations consisted of 132 seconds repetitions using single pointing mode, sparse spatial sampling and high spectral resolution (FWHM$\sim$0.048cm$^{?1}$). The data were processed using the standard pipeline described in the Observers Manual (SPIRE Observers Manual 2012) and Fulton et al. (2008).

The interferograms were cosmic ray, temperature and time-domain phase corrected. The repetitions were then averaged and Fourier transformed into the spectral domain. By taking the inverse transform of the observed interferogram we can restore the original source spectrum. Although IC342 is a grand-design spiral, it has been found that a significant fraction of the CO J=1 $\rightarrow$ 0 and J=2 $\rightarrow$1 emission arises from a $\approx$20"$\times$13" central region (e.g. Eckart et al., 1990) which is well matched to the size of the FTS SSW and SLW beams.

The (uncorrected) FTS spectrum of IC342 is shown in Figure 4.1 in red. Blue asterisks indicate the peak fluxes taken from the SPIRE photometric images (these are quoted in fluxes/beam, (Dale et al., 2012)). The prominent mismatch between SLW and SSW is the result of two effects: the variations of the FTS spectral response with

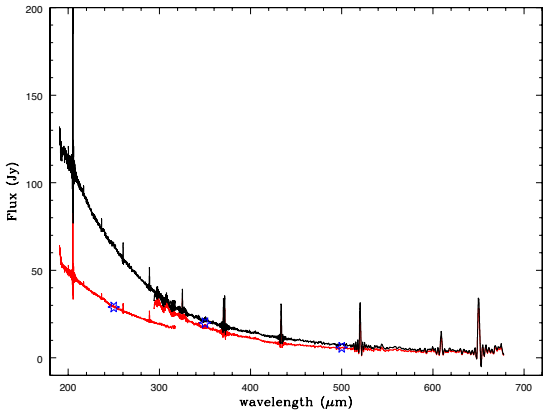


Figure 4.1: The uncorrected spectrum (in red) showing a mismatch between SSW and SLW due to the difference in beam sizes and source extent. The corrected spectrum (assuming a 19 arcsec FWHM Gaussian distribution for the source size is shown in black. Blue asterisks denote peak fluxes (flux/beam) taken from the SPIRE images (priv. comm. R. Kennicutt)

frequency and, of the way the beam couples to sources of varying extent. The first effect has been modeled with a combination of Hermite-Gauss (HG) polynomials, which follow the form of the expected native feedhorn modes (Ferlet et al., in prep.). The efficiency with which the beam couples to a given source has been estimated using observations of Neptune and Uranus to establish the point source response, and a model of the flux from the Herschel telescope itself to establish the fully extended response. IC342 fills the SSW beam (low frequency) but can be considered a point source in the SLW beam (high frequency). Following the method described above and detailed in Fletcher et al. (2012) and Wu et al. (2013) we infer the FWHM size of the source to be ∼19 arcsec. The corrected FTS spectrum (corresponding to a 19" FWHM core) is shown in Figure 4.1 (black).

We measured line fluxes from the calibrated unapodized spectrum using our own IDL-based routines. In brief, from each of the SSW (SLW) spectrum we first fit the underlying continuum which must be removed before fitting the lines. After subtracting the continuum fit each line is fitted separately using a sinc function with central frequency, line width, amplitude and residual value (in most cases this equals zero since we have removed the continuum) as free parameters. We also measured fluxes using another IDL based line fitting program called SLIDE developed by A. Rykala (priv. comm.). The integrated line fluxes derived from the two independent

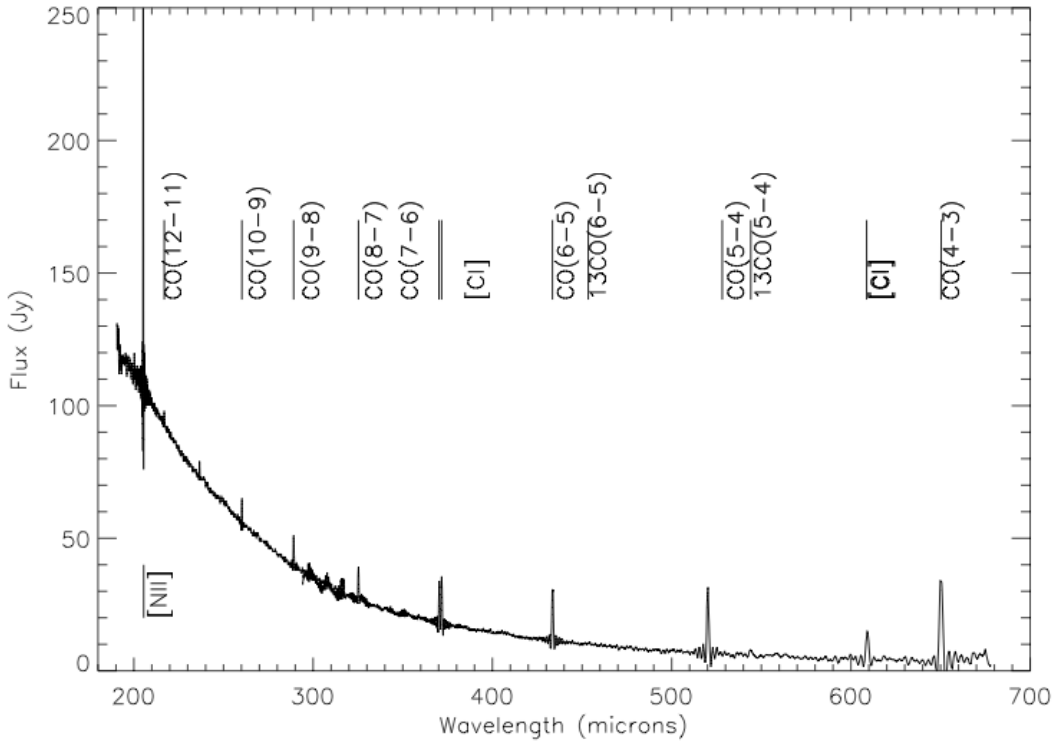


Figure 4.2: The FTS spectrum of IC342 with the atomic, ionic and molecular lines identified.

methods agree very well. We note that the errors reported in Table 4.1 are $1\sigma$ uncertainties estimated from the fitting method. Figure 4.2 shows the atomic, ionic and molecular lines identified in the IC342 FTS spectrum.

## 4.4 Modeling the CO lines

The properties of the molecular gas in IC342 have been studied through low-J (J<6) CO emission lines by, amongst others, Eckart et al. (1990), Harris et al. (1991) and (Bayet et al., 2006). Extensive modeling of the low-J CO lines by Bayet et al. (2006) revealed gas densities of the order of a few $\times 10^3$ cm$^{-3}$ and gas (kinetic) temperatures around 40 K, for the low density, low temperature component. Through spatially resolved kinematics of the CO(6–5) line, Harris et al. (1991) suggest that the warm gas is not originating from the same location as the CO low-J lines. With the SPIRE-FTS we can now sample a large part of the $^{12}$CO-ladder from J=4–3 to J=13-12. These high-J CO lines, inaccessible from the ground, supplemented with detections of lower-J CO transitions allow us to carry out a full investigation of the physical conditions of the molecular gas in IC342.

Table 4.1: Measured fluxes of detected emission lines

| Species | Transition | Wavelength (rest, $\mu$m) | $F_\lambda$ ($10^{-17}$ W m$^{-2}$) | $F_\nu d\upsilon$ ($10^3$ Jy km s$^{-1}$) |
|---|---|---|---|---|
| $^{12}$CO | J=4–3 | 650.245 | 40.15±0.81[1] | 26.38 |
| $^{12}$CO | J=5–4 | 576.268 | 35.62±0.75 | 18.73 |
| $^{12}$CO | J=6–5 | 433.556 | 31.23±0.69 | 13.68 |
| $^{12}$CO | J=7–6 | 371.650 | 24.23±0.87 | 9.10 |
| $^{12}$CO | J=8–7 | 325.225 | 19.27±0.82 | 6.33 |
| $^{12}$CO | J=9–8 | 289.118 | 17.12±1.14 | 5.00 |
| $^{12}$CO | J=10–9 | 260.238 | 12.52±0.73 | 3.30 |
| $^{12}$CO | J=11–10 | 236.611 | 8.912±0.95 | 2.13 |
| $^{12}$CO | J=12–11 | 216.925 | 6.342±0.82 | 1.39 |
| $^{12}$CO | J=13–12 | 200.271 | 3.59±1.24 | 0.72 |
| $^{12}$CO | J=1–0 | 2606.869 | 2.25±0.21[2] | 5.93 |
| $^{12}$CO | J=2–1 | 1303.434 | 8.85±0.76[2] | 11.66 |
| $^{12}$CO | J=3–2 | 868.956 | 22.51±1.19[2] | 19.77 |
| $^{13}$CO | J=5-4 | 544.156 | 3.31±0.85 | 1.82 |
| $^{13}$CO | J=6-5 | 453.494 | 1.18±0.40 | 0.54 |
| $[NII]$ | $^3P_1 \rightarrow ^3P_0$ | 205.226 | 117.65±8.2 | 54.64 |
| $[CI]$ | $^3P_1 \rightarrow ^3P_0$ | 370.466 | 21.70±0.8 | 8.124 |
| $[CI]$ | $^3P_2 \rightarrow ^3P_1$ | 608.812 | 9.04±0.9 | 5.57 |

[1]Quoted errors represent $1\sigma$ errors from the line fitting procedure and do not include (e.g.) instrumental uncertainties.

[2]Values from Bayet et al (2006) and references therein, data taken with beamsize 21.9".

For the present investigation of the physical conditions of the molecular gas we used the non-LTE radiative transfer code RADEX (van der Tak et al., 2007) to compute CO intensities (from J=1–0 to J=13–12) for a large grid of temperatures $T_{kin}$, density n($H_2$), column density ($N_{CO}$) and source size. We use the uniform expanding sphere approximation and a 2.73 K blackbody to represent the cosmic microwave background (CMB). As discussed by Kamenetzky et al. (2012) the choice of background does not affect the resulting kinetic temperature of the model component(s). The code starts off with the optically thin case and generates level populations. The process continues until a stable self-consistent solution is found where the optical depth of the lines remain stable from one iteration to the next. Using the code we have searched a large grid of parameters in $T_{kin}$: 10 – 3000 K, n($H_2$) : $10^2$ – $10^8$ $cm^{-3}$, $N_{CO}$ : $10^{15}$–$10^{24}$ $cm^{-2}$ and, source size : 0–1000 $arcsec^2$. During the search procedure we reject those models where: (a) the optical depth of the low-J lines modeled is outside the range -0.1 $< \tau <$100 (e.g. van der Tak et al., 2007; Kamenetzky et al., 2012) and (b) $M_{gas}$ becomes larger than $M_{dyn}$, the dynamical mass of the galaxy. We note that $M_{gas}$ is defined as:

$$M_{\rm gas} = \Omega D_{\rm A}^2 N_{\rm CO} \times \frac{\mu m_{H_2}}{x_{\rm CO}} \tag{4.1}$$

where, $x_{CO}$=$3\times10^{-4}$ is the relative CO/$H_2$ abundance, $D_A$ is the angular distance diameter in $cm^2$, $\Omega$ is the angular source size and $\Omega\times$ $D_A^2$ is the source size. The mean molecular weight, $\mu$ =1.4, is in units of $m_{H_2}$. The source size remains the same for all transitions. We finally assume a line width of 54 km $s^{-1}$ (Bayet et al., 2006).

The parameter space is searched using the nested sampling routine *Multinest* (Feroz & Hobson, 2008). In brief, *Multinest* is a Bayesian inference tool for model selection and parameter estimation. It is based on the Monte Carlo technique of nested sampling (Skilling, 2004), which can evaluate the Bayesian evidence (useful for model selection) and sample from posterior distributions with (often an unknown number of) multiple modes and/or degeneracies between parameters. The posterior distribution $P_r(M|x)$ gives the probability of the model parameters (M) given a set of measurements x. Using Bayes theorem, the posterior can be expressed as:

$$P_r(M|x) = \frac{P_r(M)P_r(x|M)}{P_r(x)} \tag{4.2}$$

Table 4.2: Parameters of the single component model

| | $Log_{10}T_{kin}$ | $Log_{10}n(H_2)$ | $Log_{10}N_{CO}$ | $Log_{10}M_{gas}$ | $\Omega$ |
|---|---|---|---|---|---|
| | K | $cm^{-3}$ | $cm^{-2}$ | $M_\odot$ | arcsec |
| mean | 2.70 | 2.5 | 18.5 | 7.1 | 227 |
| std dev | 0.02 | 0.2 | 0.2 | 0.2 | 31 |
| ML | 2.69 | 2.7 | 18.2 | 6.8 | 224 |

where $P_r(M)$ is the prior probability that a set of parameters is either physical or unphysical and is set to the grid ranges described above. $P_r(M|x)$ is the likelihood of reproducing the observational measurements with a RADEX SLED, given a specific set of model parameters. We use the standard Gaussian likelihood measure to calculate the likelihood. $P_r(x)$ is the normalisation parameter or Bayesian Evidence which is used for model selection. In order to find the posterior distribution for one parameter, e.g. $T_{kin}$, we need to marginalise over all other parameters to find the probability $P(T_{kin})$.

## 4.5 Results and Discussion

We use RADEX and the nested sampling routine *Multinest*, described above, to model the CO line intensities. All the CO lines reported in Table 4.1 are used in the analysis supplemented with low-J CO lines reported in the literature. For the total uncertainty, we take the $1\sigma$ statistical uncertainty in the total integrated intensity from the line fitting procedure and add 10% calibration error in quadrature. We first consider a one-component model (assuming a uniform temperature for all transitions). The highest likelihood model is shown in Figure 4.3. The resulting 'posterior' distributions for each parameter marginalised over the other parameters used in the model are shown in Figure 4.4. In Table 4.2 we report the mean, standard deviation and maximum likelihood (ML) values for the $T_{kin}$, $n(H_2)$, $N(_{CO})$ and $\Omega$.

The deviation of the low J (J$\leq$3) transitions from our highest likelihood model indicates the presence of a second (possibly colder) gas component primarily responsible for the low-J lines. We thus consider a two-component model to fit the CO-SLED following the same procedure as before. In our two component model we allow the three RADEX parameters $T_{kin}$, $n(H_2)$ and $N_{CO}$ to vary independently for both components, giving us 6 free parameters to search over. We restrict

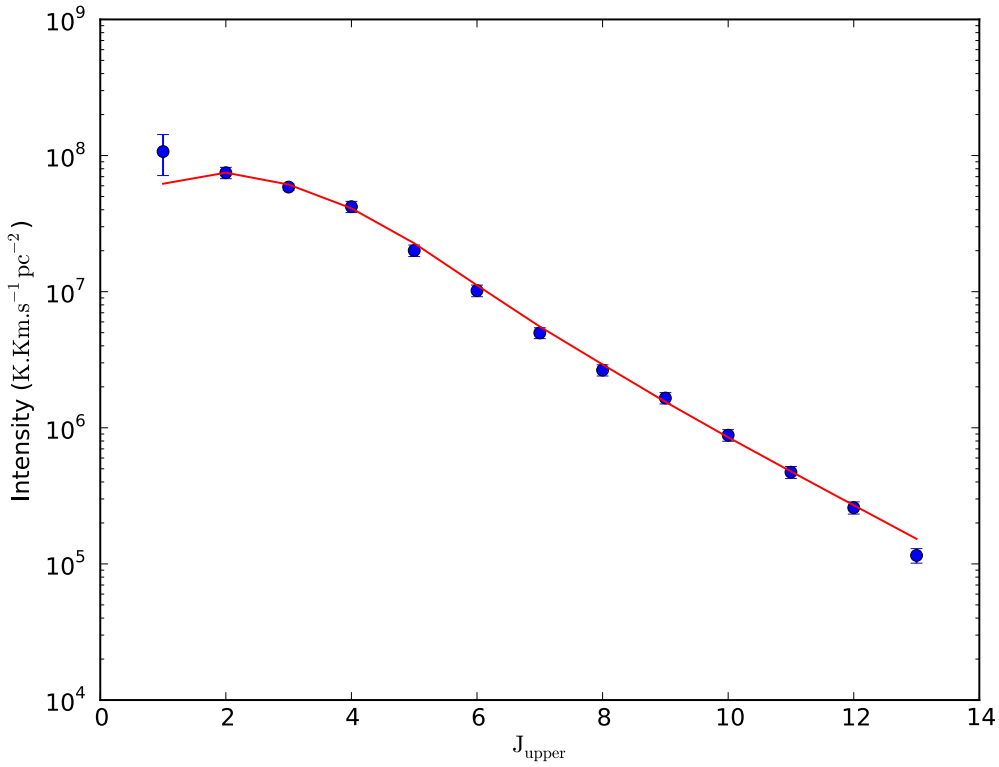


Figure 4.3: The highest likelihood model (red line) is compared to our data (blue points) and associated errors.

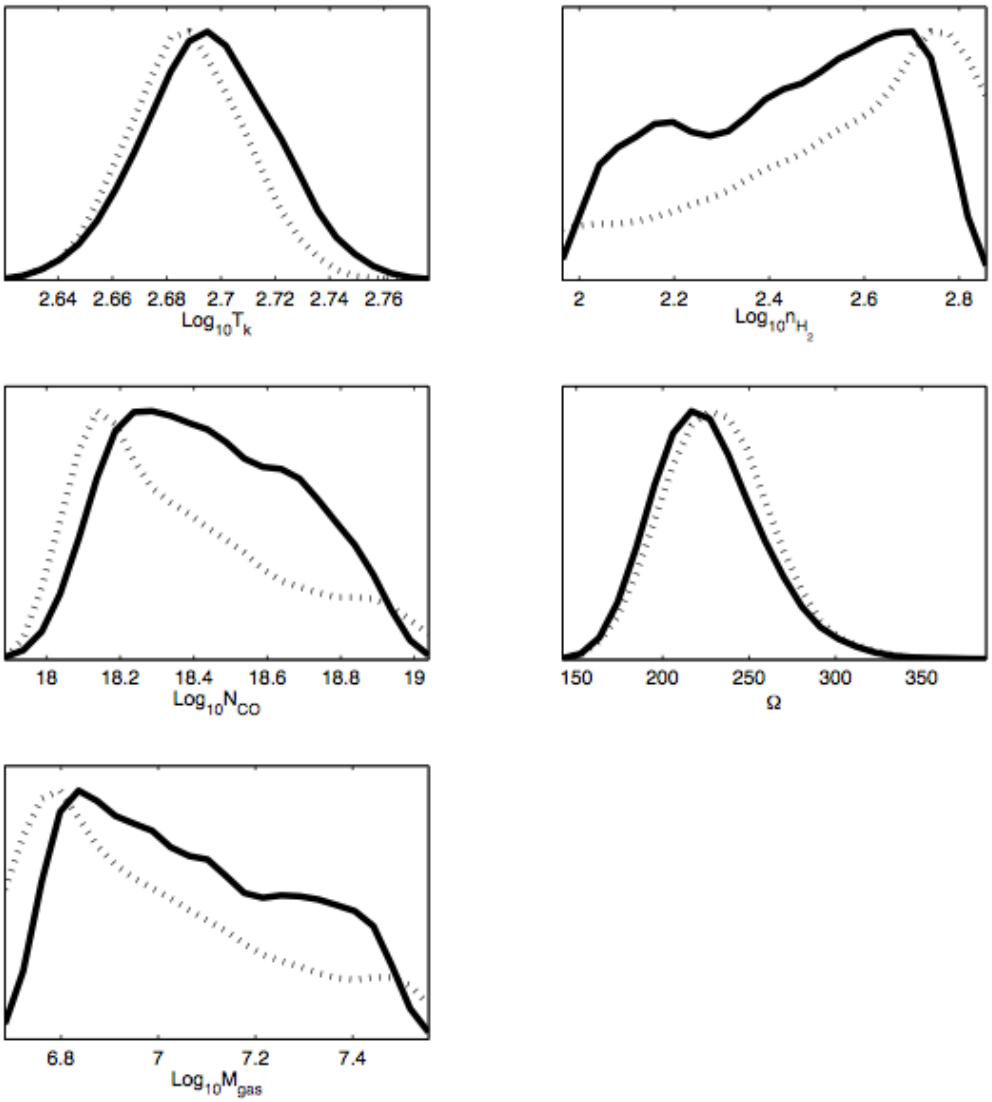


Figure 4.4: Maximum Likelihood distributions of kinetic temperature, $H_2$ density, CO column density, $M_{gas}$ and $\Omega$ source extent. The dotted lines represents the mean distributions of the values.

the temperatures of the two components to be in the range 10< $T_k$ <200K and 200< $T_k$ <3000, thus, defining a warm and cold component. We fix the source size to the size of the FTS beam (2140$arcsec^2$) as the cold gas is likely to fill the entire beam and, place an upper limit to the dynamical mass of the combined gas mass from both components.

Unlike previous studies that model the CO-SLED of galaxies (e.g. Kamenetzky et al., 2012) we do not restrict the components to any particular subset of CO lines. That is, both components are combined to give the overall line luminosity. Since there are few constraints on the actual source size of the emitting region from the different CO lines, we also fix the source size of both components to the size of the FTS beam (2140arcsec$^2$). In order to ensure that our results are not significantly biased by fixing the source size we have repeated our models allowing the source size to vary. Such runs have indicated that there exist a degeneracy between source size and column density which is rather difficult to break. As a result we decided to fix the source size so that we can have an independent estimate for the CO column density of the two components without concerns about possible degenerate values.

In Figure 4.5 we show the highest likelihood two-component model that provides the best fit to the available $^{12}$CO data. The low-J lines originate in the cold gas component (40 K) while the warmer gas component (437 K) is responsible for the higher-J lines. The marginalised plots for each parameter with the mean likelihood shown as dotted line are presented in Figure 4.6. In Table 4.3 we report the detailed values of the physical characteristics of the warm and cold gas.

In Figure 4.5 we show the 2-d marginalised contour plots for the three main parameters, $T_{kin}$, $\eta_{H_2}$ and $N_{CO}$. The plots (1 and 2$\sigma$ contours) show the range of parameters for the two temperature components (blue contours for the cold and red contours for the warm component). The range of temperatures for the cold component is much wider than for the warm component. Two reasons are likely contributing to this effect. First, *Multinest* is very sensitive to errors (of the CO line fluxes) and the error reported for the CO(1–0) transition is driving the fit. Second, for the two component model, we have fixed the apparent source size to the size of the FTS beam. It is very likely that the cold gas may originate in a more spatially extended area than gas at warmer temperatures. The equivalent parameters ($T_{kin}$, versus $\eta_{H_2}$ and $N_{CO}$) for the warm component show a much narrower variation.

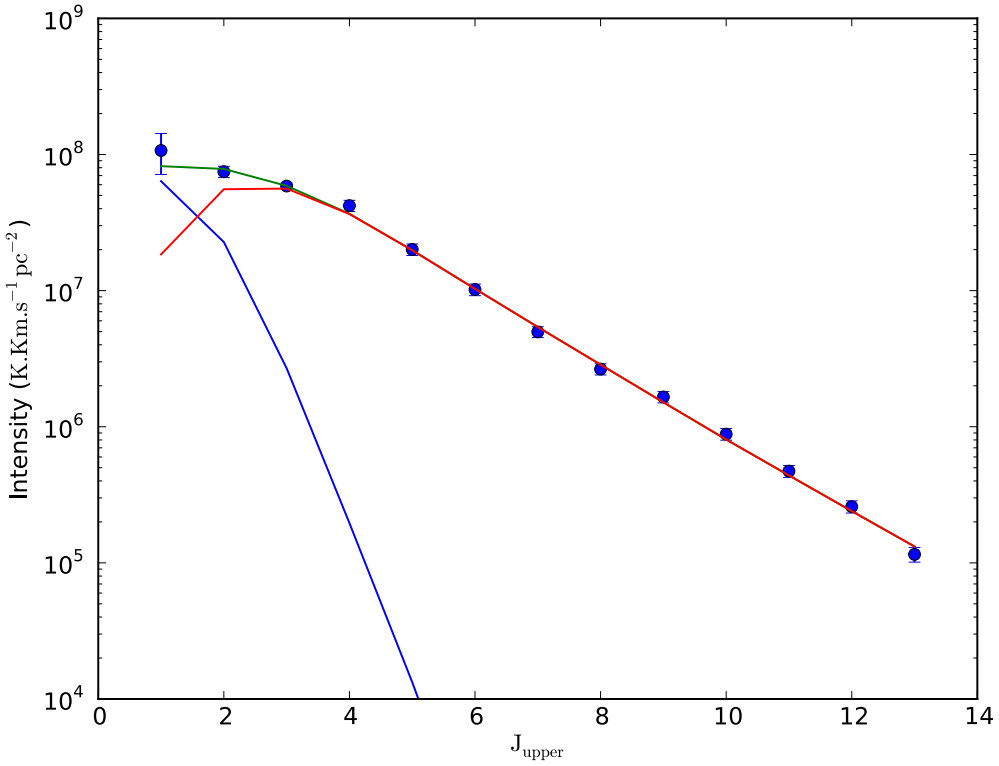


Figure 4.5: Maximum Likelihood Analysis of the $^{12}$CO Spectral Energy Distribution. The CO measurements and associated error bars are shown as blue filled circles. The blue line represents the cold component while the red line represents the warm component. The total of the two components is the green line.

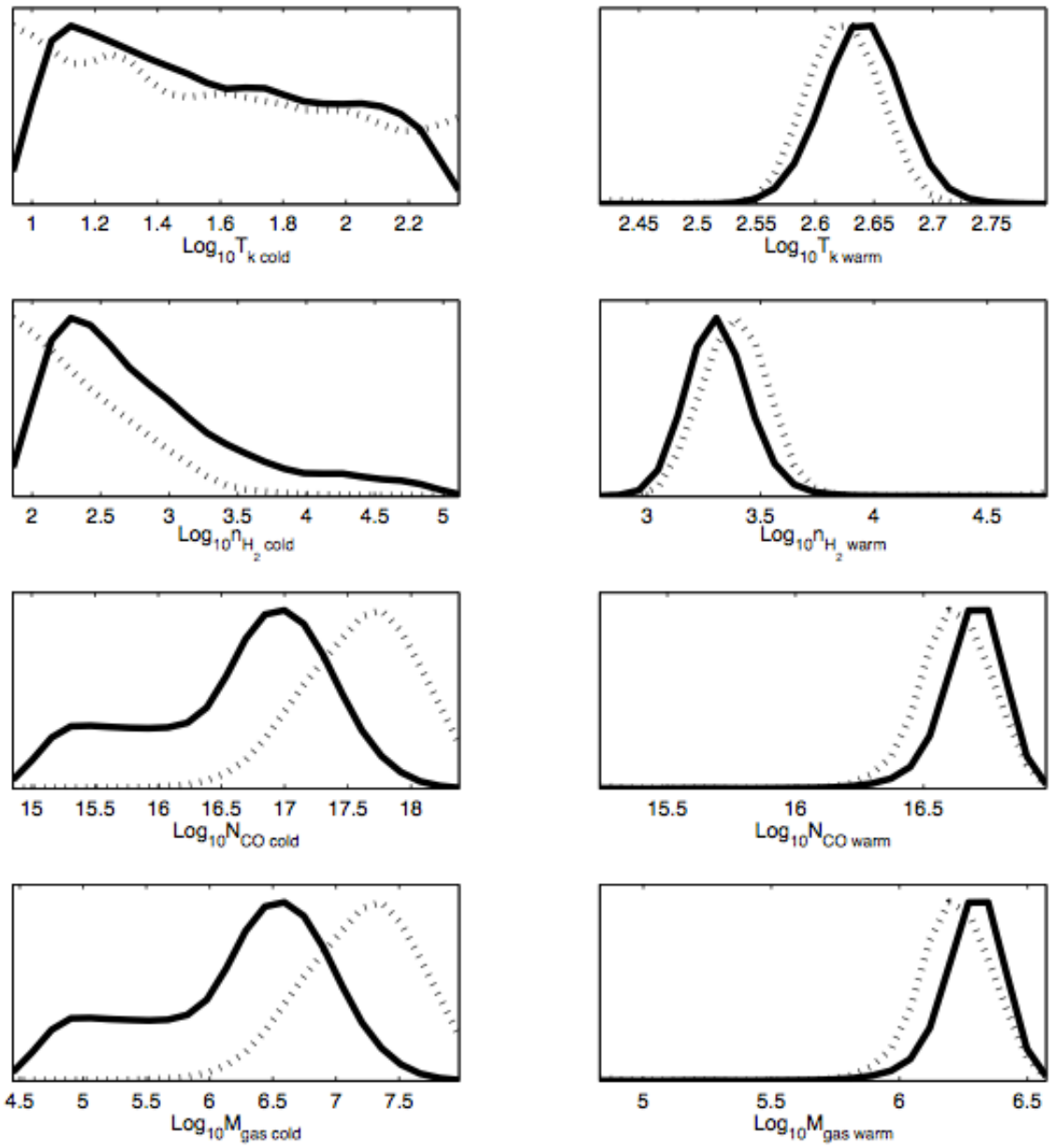


Figure 4.6: Posterior distributions of kinetic temperature, $H_2$ density, CO column density and $M_{gas}$ for the cold (left panel) and the warm (right panel) components. The dotted line represents the mean distributions of the values.

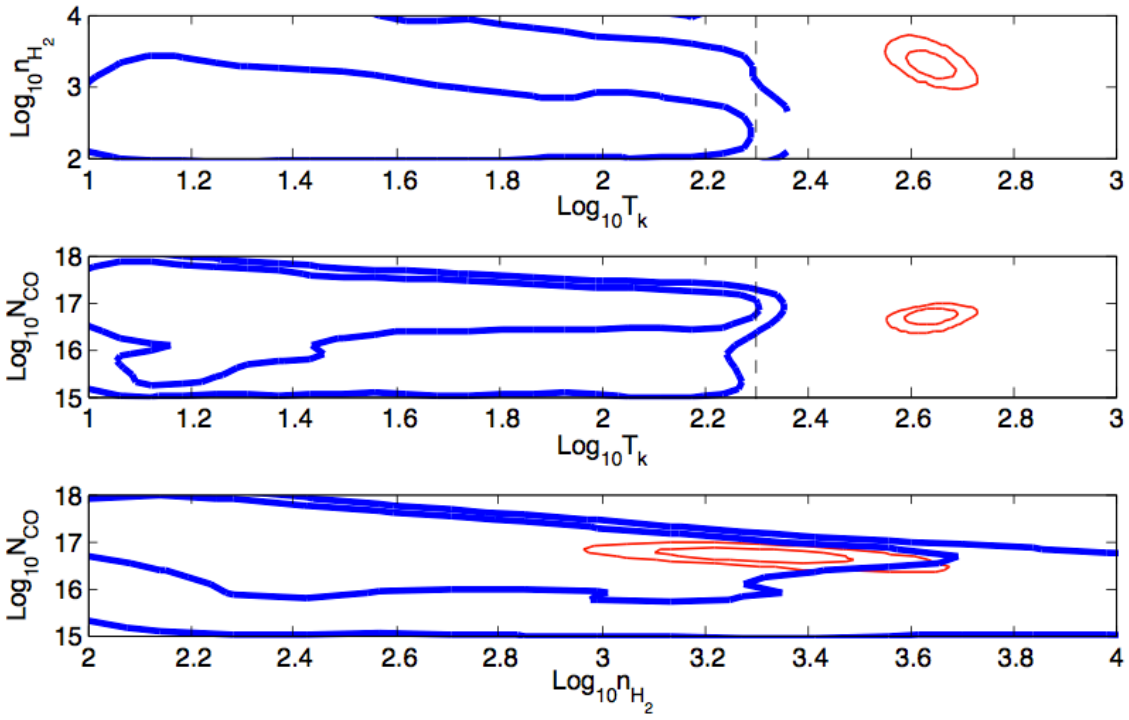


Figure 4.7: One and two sigma contour levels for the three parameters $\eta_{H_2}$, $N_{CO}$ and $T_{kin}$ for the cold (blue) and the warm (red) component.

Following equation 4.1, in Table 4.3 we report the mean, standard deviation and MLM values for the gas mass for the each of the two components considered. For the the warm component the MLM warm gas mass value is $1.9\times10^6$ $M_\odot$ with a similar mean value of $1.9\times10^6$ $M_\odot$. Rigopoulou et al. (2002) reported the detection of mid-infrared rotational $H_2$ transitions using ISO. Assuming an ortho-to-para ratio of 3 the S(1)–S(2) ratio can be used to estimate the temperature of the warm gas. For IC342 they report a warm gas temperature of 365 K which is consistent with our value of 437 K (taking into account differences in the beam sizes and calibration). Based on their sample, Rigopoulou et al. (2002) find that the warm gas mass (measured from $H_2$) accounts for 1 to 10% of the total gas mass (measured from CO) in starburst galaxies which is consistent with our findings.

The ISM in IC342 has been studied in the past using ground based measurements of (primarily low-J) CO transitions. Using LVG calculations Eckart et al. (1990) suggested $T_{kin} > 20$ K, Israel & Baas (2003) reported $T_{kin} = 100$ –150 K while Bayet et al. (2006) fit their data with a model of 40 K. All three models predict densities $n(H_2)$ of the order of 2–3 $\times10^3$ cm$^{-3}$. The $T_{kin}$ value we derive for the cold component is in good agreement with the values reported by Bayet et al. (2006) and

Table 4.3: Parameters of two-component model

| Cold Component | | | | |
|---|---|---|---|---|
| | $Log_{10}T_{kin}$ K | $Log_{10}n(H_2)$ $cm^{-3}$ | $Log_{10}N_{CO}$ $\times 10^{17}$ $cm^{-2}$ | $Log_{10}M_{gas}$ $M_\odot$ |
| mean | 1.6 | 2.8 | 16.6 | 6.2 |
| std dev | 0.4 | 0.7 | 0.7 | 0.7 |
| ML | 1.2 | 2.1 | 17.7 | 7.3 |
| Warm Component | | | | |
| mean | 2.64 | 3.32 | 16.68 | 6.28 |
| std dev | 0.03 | 0.14 | 0.13 | 0.13 |
| ML | 2.63 | 3.32 | 16.69 | 6.23 |

Eckart et al. (1990) but lower than the value reported by Israel & Baas (2003). **Our value also agrees with the T$\sim$53$\pm$1 K and the $T_{warm}\sim$443$\pm$130 K derived by Mauersberger et al. (2003) for the cold and warm gas components based on $NH_3$ observations.** The maximum likelihood value for the $H_2$ density of the cold temperature component predicted by our model is, however, lower than those reported in the literature. This is perhaps not surprising given that the cold gas component is in fact expected to have a lower density and a higher $M_{gas}$. The warm gas component however, has a higher $n(H_2)$ density, and a lower column density $N_{CO}$. Based on the the lowest J $^{13}$CO lines, Meier et al. (2000); Meier & Turner (2005) inferred gas temperatures of 10–20 K and densities of $\approx 10^{3.5} cm^{-3}$. It is thus likely that there are multiple components to the molecular gas. Denser gas with $n(H_2) > 10^6$ is also present in IC342 as traced through CS and $HC_3N$ lines (Aladro et al., 2011). This dense gas is unlikely to contribute dominantly to the observed CO emission, instead it is found in the central core of the molecular clouds in IC342 where evidently (e.g. Martín et al., 2009) is it used to sustain high star formation efficiency.

## 4.6 The Origin of the Warm CO molecular Gas

A number of mechanisms can heat the CO molecular gas, including UV starlight in photodissociation regions (PDRs), X-ray heated gas (possibly associated with the presence of an AGN) in X-ray dominated regions (XDRs), mechanical heating (e.g.

turbulence dissipation), powerful shocks and heating by cosmic rays. The absence of a strong AGN in IC342 together with the shape of the CO SLED can probably rule out XDRs as the origin of the warm CO emission. In the case of XDR dominance the CO SLED becomes flat at higher-J CO transitions (e.g. Mrk 231,van der Werf et al. (2010)) which is clearly not the case in IC342.

A cosmic-ray ionisation rate of $\sim 5\times10^{-17}$ s$^{-1}$ (e.g. de Jong et al., 1980) is sufficient to heat the gas to about $\sim$10 K. For an n($H_2$) density of $6\times10^2$ cm$^{-3}$, CO column density of $3.16\times10^{17}$ cm$^{-3}$ and an excitation temperature of 35 K the cosmic-ray ionisation rate is $\approx 10^{-15}$ s$^{-1}$ which is about 20 times higher compared to the standard rate. As Eckart et al. (1990) argue this is unlikely to be the case, especially in the nuclear region of IC342 as it would require a very strong 5 GHz non thermal radio component which is not observed. Therefore, although cosmic rays may contribute towards heating low temperature $\sim$10 K gas it is unlikely to be the dominant mechanism of CO heating.

Recent/ongoing star-forming activity in IC342 (e.g. Boker et al., 1997; Meier & Turner, 2005) has resulted in an increased number of OB stars and therefore, in a more intense UV radiation field in this galaxy. The UV radiation heats up the surfaces of the molecular clouds by means of photoelectric heating and heating through far-UV pumping of $H_2$ (e.g. Tielens & Hollenbach, 1985). PDRs form in the outer layer of the clouds which are responsible for CO emission from warm/hot gas. Using the grid of PDR models presented in Kaufman et al. (1999, 2006) we investigate predictions for the various $^{12}$CO lines. The rest of this section deals with models of the $^{12}$CO lines although for simplicity we refer to them as CO lines. The PDR models cover a wide range in $H_2$ density (10–$10^7$ cm$^{-3}$) and $G_0$, the incident UV flux. The parameter $G_0$, called the Habing interstellar radiation field, is defined as FUV flux $=1.3\times10^{-4}\times G_0$ erg cm$^{-2}$ s$^{-1}$ sr$^{-1}$. The FUV flux is related to the FIR flux via the relation FIR flux=2$\times$ FUV. We note, however, that in our investigations of the PDR models we do not seek to determine the parameters that best fit the observed line ratios. Instead, in addition to the relative CO line intensities, we use the physical conditions determined from our LVG modeling as priors for PDR models. In our search we do not only consider the ML (or mean) value, instead, we make use of our likelihood analysis and take into account a larger parameter grid (see Figure 4.5). With the above caveats in mind, we have used the

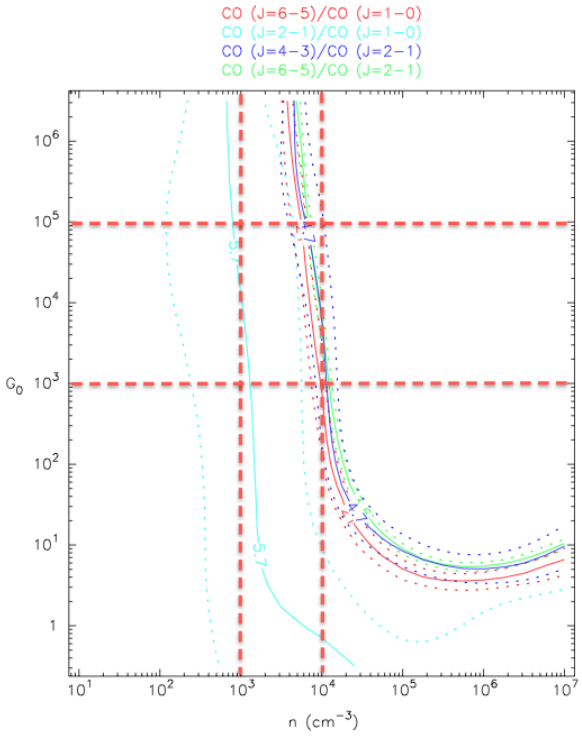


Figure 4.8: CO line ratios as a function of incident FUV flux ($G_0$ and and density $\eta$ for the PDR models. The vertical red dotted lines indicate the range of values for $\eta$ predicted by the RADEX models while the horizontal lines indicate the range of $G_0$ values required to match the observed CO line ratios.

online PDR Toolbox program * (Kaufman et al., 2006; Pound & Wolfire, 2008) to investigate the range of PDR parameters that would best match our CO line ratios. The PDR Toolbox program calculates the best values of $G_0$ and cloud density $\eta$ for a given set of spectral line intensities. A description of the PDR models used for these tools can be found in Kaufman et al. (1999). For a given set of gas phase elemental abundances and grain properties, each model is described by a constant H nucleus density, n, and incident far-ultraviolet intensity $G_0$. The models solve for the equilibrium chemistry, thermal balance, and radiation transfer through a PDR layer. We used the line intensities quoted in Table 4.1 supplemented by those from the literature. The FTS lines have been measured assuming a FWHM of $19''$ as discussed in Section 4.3. Figure 4.8 shows line ratios for various CO transitions. Assuming the range of $n(H_2)$ values determined from the likelihood analysis ($10^3$–$10^4$ cm$^{-3}$) $G_0$ should be in the range$10^3$ $<G_0$ $<10^5$ to match the observed CO line ratios. In fact the best-fit PDR models predicts $n(H_2) = 5.62\times10^3$ cm$^{-3}$ and $G_0$=$3.16\times10^5$. Although at this stage a second mechanism (e.g. cosmic ray heating of the coldest gas) contributing to the CO lines cannot be excluded it is very likely that PDRs play an important, perhaps dominant, mechanism for the CO emission we have detected. Our measured line ratios can be matched with PDR models with $n(H_2)$ very close to the values found by the maximum likelihood analysis. Based on the models presented we conclude that a significant fraction of the CO emission in IC342 originates in PDRs.

## 4.7 Far Infrared Atomic Fine Structure Lines

As discussed in section 4.3 (and presented in Table 4.1) we detect three fine structure lines, the two [CI] 370 and 608 $\mu$m and [NII] 205 $\mu$m lines. Prior to Herschel only a handful of [CI] extragalactic measurements had been carried out from the ground. The [NII] 205 $\mu$m line emission, in particular, is very difficult to observe from the ground because it lies near the long-wavelength cutoff for stressed Ga:Ge photoconductors, strongly limiting the sensitivity of spectrometers in this regime. Oberst et al. (2006) reported the detection of [NII] 205 $\mu$m towards the Carina galactic star forming region. With SPIRE-FTS we now have additional [NII] 205 $\mu$m measurements in M82 (Kamenetzky et al., 2012; Panuzzo et al., 2010), Arp220

*http://dustem.astro.umd.edu/pdrt/

(Rangwala et al., 2011).

The ratio of the [CI] lines can be used to estimate the physical properties of the atomic [CI] gas that arises predominantly in the CO dissociating regions. For IC342 we derive a I([CI]370$\mu$m)/I([CI]608$\mu$m) ratio of 1.45±0.3 which is close to the value of 1.62 ±0.2 using the measurements reported for M82 in Kamenetzky et al. (2012) but higher than the value of 1.15±0.3 derived for Arp220 by Rangwala et al. (2011). We use the escape probability radiative transfer models presented in Stutzki et al. (1997) and their Figure 2, to infer the [CI] column density $N_{[CI]}$, $T_{kin}$ and hydrogen column density n($H_2$). When expressed in line temperature units (K km s$^{-1}$) the [CI](J=2→1)/[CI](J=1→0) ratio for IC342 becomes 0.54. For this observed line ratio, a brightness temperature of ≈ 15 K and assuming a line width of 54 km s$^{-1}$ we estimate a column density N(C)=2.1×10$^{17}$ cm$^{-2}$ and $H_2$ density of 5.4×10$^3$ cm$^{-3}$. These values are indicative of an excitation temperature 40<$T_{kin}$<60 K. The range of the inferred [CI] excitation temperature is slightly higher than the one derived from cold (low-J) CO (mean value of 36 K) although well within the uncertainties involved. It is, thus likely that that [CI] may originate in the same (cold) molecular gas. The column density ratio $N_C/N_{CO}$ is 0.66 a value similar to that derived for M82 (e.g. Stutzki et al., 1997) but lower than the value of 1 reported for Arp 220 by Rangwala et al. (2011). It has been suggested (e.g.Wilson (1997) but see also discussion in Rangwala et al. (2011)) that [CI] emission is stronger for more luminous systems or those harboring intense starbursts or AGN.

Israel (2005) examined the [CI](2→1)/$^{12}CO(4-3)$ and [CI](2→1)/$^{13}CO(2-1)$ line ratios for a sample of quiescent, starburst and active galaxies. He found that in the majority of galaxies the [CI](2→1)/$^{13}CO(2-1)$ ratio is ≥2. Lower ratios are expected in high-UV environments with high column densities where the majority of neutral C will be locked up in CO. For IC342 the [CI](2→1)/$^{13}CO(2-1)$ value of 1.2 (Israel, 2005) is smaller than the values found for the majority of strong starbursts or luminous systems. The line ratio together with the derived C column density of 2.1×10$^{17}$ cm$^{-2}$ and the fact that [CI] is weaker than CO emission in IC342 could be be indicative of [CI] arising in a thin layer on the outside of the CO emitting molecular clouds consistent with PDRs playing an important role.

The [NII]205 $\mu$m and [CII]158 $\mu$m lines have nearly identical critical densities for excitation in ionised gas regions. Their line ratio is thus insensitive to the hardness

of the stellar radiation field (since the photon energies required to ionize each species to the next ionisation states are similar) and is only a function of the [NII]205/[CII] abundance ratio. The ratio [NII]/[CII] can, thus, give us an estimate of the fraction of [CII] arising in ionised gas for a given value of the ionised gas density. The [NII] 122/205 $\mu$m line ratio can be used to probe the density of the ionized gas.

Using [CII] and [NII]122 $\mu$m measurements of IC342 from the ISO-LWS archive (but also in Brauher et al. (2008) for [CII]) we infer a [NII] 122/205 $\mu$m line ratio of 3.2 and a [CII]/[NII]205 ratio of 17$\pm$2.5. We compare these ratios to the model line intensity ratio as a function of ionized gas density in Figure 2 of Oberst et al. (2006). Their [NII] 122/205 model line ratios range from 2.5 to 4.3, therefore, our estimated [NII] 122/205 $\mu$m line ratio of 3.2 indicates a density of $\sim 200$ cm$^{-3}$. For this density, if both lines arise from the ionized gas the expected [CII]/[NII] 205 ratio would have been $\sim$3. Given our estimate of 17$\pm$2.5 we infer that the contributions from ionised gas to [CII] is 17-20% with the remaining originating in warm gas PDRs. This result lends further support to the suggestion that PDRs must play an important role in IC342.

## 4.8 CO Ladder: an insight into the excitation of molecular gas of galaxies

The most widely used method to investigate the properties of the molecular gas of galaxies near and far, is through measurements of CO transitions. For nearby galaxies, a suite of low-J CO transitions are readily accessible from the ground and have been used to measure the excitation of the molecular gas (e.g. Boone et al., 2011). For high-redshift galaxies however, we can only obtain a handful of CO transitions, depending on redshift (e.g. Solomon & Vanden Bout, 2005; Wagg et al., 2010), hence the resulting CO-SLEDs are very sparsely sampled. The situation is likely to change with the availability of the full ALMA array.

With the SPIRE-FTS we can now measure a number of high-J CO transitions (SPIRE-FTS can access up to J=13 although higher CO transitions up to J=20 can be observed with PACS) and when combined with ground based measurements we can determine the CO-SLED to much greater accuracy. As Kamenetzky et al. (2012) point out, the diagnostic tools to model the high-J CO lines are still under

development, however, we have already began to exploit the new information available. Extremely high CO line excitation implies the presence of an AGN and strong heating by XDRs or shocks. In star forming regions PDRs play a crucial role by heating the dust efficiently. The CO-SLED of Mrk 231 has been modelled as the result of contributions from PDRs and XDRs, the latter to account for the emission of the high-J lines (van der Werf et al., 2010). In lower luminosity AGN signatures of XDR heating are present close to the nucleus on small scales while in the starburst regions CO excitation originates in PDRs (e.g. Spinoglio et al., 2012). In Starbursts, such as M82, detailed analysis of the CO-SLED revealed that although PDRs could play a role they cannot provide high enough densities to match the observed CO line emission. Kamenetzky et al. (2012) have therefore concluded that mechanical energy (turbulent motions, shocks, cosmic ray heating) is likely to play a large role in the heating of the galaxy. A similar conclusion has been reached by Rangwala et al. (2011) for Arp220.

In the case of IC342, we have shown that the CO-SLED can be adequately explained with a combination of a cold (T$\sim$40 K), low density component to explain the low-J transitions (with most contribution from J=1-0) and a warmer component (T$\sim$440 K) necessary to account for the high-J transitions. **There is convincing evidence that CO excitation in this quiescent star-forming galaxy can be provided by PDRs without invoking a need for extra heating mechanisms. Thus, IC342 may be used as a template for explaining the CO-SLED of high-redshift main-sequence galaxies** that do not appear to be merger-driven (e.g. Elbaz et al., 2011).

## 4.9 Conclusions

We have presented spectroscopic observations of IC342 covering the wavelength range 194-671 $\mu$m carried out with the SPIRE-FTS on board Herschel. We have detected a number of $^{12}$CO, $^{13}$CO molecular, [CI] atomic and [NII] ionic lines. We have used the radiative transfer code RADEX coupled with *Multinest* to model the $^{12}$CO SLED: the modeling procedure has revealed two gas components, a cold temperature component T$\sim$35 K and, a much warmer component with T$\sim$400 K. Our new SPIRE-FTS data and in particular the high-J $^{12}$CO lines have allowed us to constrain the physical properties of the warm gas component. Based on the ob-

served CO line ratios and published models of PDRs we argue that the CO emission originates in PDRs although additional contributions from e.g. cosmic ray heating cannot be excluded. Using the observed atomic [CI] line ratio we discuss the origin of the [CI] and suggest that it arises in a thin layer on the outside of the CO emitting molecular clouds. This claim is further supported by the fact that the [CI] emission in IC342 is much weaker than the CO(4–3). Using [CII] 158 $\mu$m measurements from ISO (Brauher et al., 2008) and the [NII] 205 $\mu$measurement presented here we infer that up to 70% of [CII] arises in warm PDRs on the surface of molecular gas clouds. Such detailed studies of the properties of the atomic, molecular and ionised gas in nearby galaxies have only recently become available thanks to the capabilities of Herschel. These studies are however, very important since they provide templates for understanding the physics of the ISM of high redshift bright submillimetre galaxies that are currently impossible to study at the same level of detail. ALMA however, in its full potential will be able to access the CO-ladder of high redshift galaxies and extend our understanding of the properties of the ISM in extreme environments.

## Acknowledgments

GEM acknowledges support from the John Fell Oxford University Press (OUP) Research Fund. DAR acknowledges useful discussions with Rob Kennicutt, Daniela Calzetti and Chad Engelbracht and, in particular, the availability of SPIRE data for IC342 ahead of publication. SPIRE has been developed by a consortium of institutes led by Cardiff Univ. (UK) and including: Univ. Leth- bridge (Canada); NAOC (China); CEA, LAM (France); IFSI, Univ. Padua (Italy); IAC (Spain); Stockholm Observatory (Sweden); Imperial College London, RAL, UCL-MSSL, UKATC, Univ. Sussex (UK); and Caltech, JPL, NHSC, Univ. Colorado (USA). This development has been supported by na- tional funding agencies: CSA (Canada); NAOC (China); CEA, CNES, CNRS (France); ASI (Italy); MCINN (Spain); SNSB (Sweden); STFC, UKSA (UK); and NASA (USA).

# Bibliography

Aladro R., Martín-Pintado J., Martín S., Mauersberger R., Bayet E. (2011). CS, $HC_3N$, and $CH_3CCH$ multi-line analyses toward starburst galaxies. The evolution of cloud structures in the central regions of galaxies. A&A, 525:A89, 1009.1831. Cited on 125

Bayet E., Gerin M., Phillips T. G., Contursi A. (2004). The submillimeter C and CO lines in Henize 2-10 and NGC 253. A&A, 427:45–59, astro-ph/0407223. Cited on 112

Bayet E., Gerin M., Phillips T. G., Contursi A. (2006). A survey of submillimeter C and CO lines in nearby galaxies. A&A, 460:467–485, astro-ph/0611340. Cited on 112, 115, 117, 124

Becklin E. E., Gatley I., Matthews K., Neugebauer G., Sellgren K., Werner M. W., Wynn-Williams C. G. (1980). Infrared emission and star formation in the central regions of the galaxy IC 342. ApJ, 236:441–447. Cited on 112

Boker T., Forster-Schreiber N. M., Genzel R. (1997). Near-Infrared Imaging Spectroscopy of IC 342: Evolution of a Bar-Driven Central Starburst. AJ, 114:1883. Cited on 126

Boone F., et al. (2011). High-resolution mapping of the physical conditions in two nearby active galaxies based on $^{12}$CO(1-0), (2-1), and (3-2) lines. A&A, 525:A18, 1010.2270. Cited on 130

Brauher J. R., Dale D. A., Helou G. (2008). A Compendium of Far-Infrared Line and Continuum Emission for 227 Galaxies Observed by the Infrared Space Observatory. ApJS, 178:280–301, 0805.2930. Cited on 130, 132

Buettgenbach T. H., Keene J., Phillips T. G., Walker C. K. (1992). Submillimeter detection of extragalactic C I emission - IC 342. ApJ, 397:L15–L17. Cited on 112

Dale D. A., et al. (2012). Herschel Far-infrared and Submillimeter Photometry for the KINGFISH Sample of nearby Galaxies. ApJ, 745:95, 1112.1093. Cited on 112, 113

de Jong T., Boland W., Dalgarno A. (1980). Hydrostatic models of molecular clouds. A&A, 91:68–84. Cited on 126

Eckart A., Downes D., Genzel R., Harris A. I., Jaffe D. T., Wild W. (1990). Warm gas and spatial variations of molecular excitation in the nuclear region of IC 342. ApJ, 348:434–447. Cited on 113, 115, 124, 125, 126

Elbaz D., et al. (2011). GOODS-Herschel: an infrared main sequence for star-forming galaxies. A&A, 533:A119, 1105.2537. Cited on 131

Feroz F., Hobson M. P. (2008). Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. MNRAS, 384:449–463, 0704.3704. Cited on 110, 117

Fletcher L. N., et al. (2012). Sub-millimetre spectroscopy of Saturn's trace gases from Herschel/SPIRE. A&A, 539:A44. Cited on 114

Fulton T. R., et al. (2008). The data processing pipeline for the Herschel/SPIRE imaging Fourier Transform Spectrometer. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7010 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Cited on 113

Gerin M., Phillips T. G. (2000). Atomic Carbon in Galaxies. ApJ, 537:644–653, astro-ph/0003252. Cited on 111

Griffin M. J., et al. (2010). The Herschel-SPIRE instrument and its in-flight performance. A&A, 518:L3, 1005.5123. Cited on 113

Harris A. I., Stutzki J., Graf U. U., Russell A. P. G., Genzel R., Hills R. E. (1991). First observations of the CO J = 6-5 transition in starburst galaxies. ApJ, 382:L75–L79. Cited on 115

Harrison A., Puxley P., Russell A., Brand P. (1995). Atomic carbon and carbon monoxide in the nuclear region of NGC 253. MNRAS, 277:413–422. Cited on 112

Hollenbach D. J., Tielens A. G. G. M. (1999). Photodissociation regions in the interstellar medium of galaxies. *Reviews of Modern Physics*, 71:173–230. Cited on 111

Israel F. P., Baas F. (2003). CI and CO in the nearby spiral galaxies IC 342 and Maffei 2. A&A, 404:495–504, astro-ph/0304070. Cited on 124, 125

Israel F. P. (2005). Molecular gas in compact galaxies. A&A, 438:855–866, astro-ph/0504569. Cited on 129

Kamenetzky J., et al. (2012). Herschel-SPIRE Imaging Spectroscopy of Molecular Gas in M82. ApJ, 753:70, 1205.0006. Cited on 117, 121, 128, 129, 130, 131

Kaufman M. J., Wolfire M. G., Hollenbach D. J., Luhman M. L. (1999). Far-Infrared and Submillimeter Emission from Galactic and Extragalactic Photodissociation Regions. ApJ, 527:795–813, astro-ph/9907255. Cited on 126, 128

Kaufman M. J., Wolfire M. G., Hollenbach D. J. (2006). [Si II], [Fe II], [C II], and $H_2$ Emission from Massive Star-forming Regions. ApJ, 644:283–299. Cited on 126, 128

Malhotra S., et al. (2001). Far-Infrared Spectroscopy of Normal Galaxies: Physical Conditions in the Interstellar Medium. ApJ, 561:766–786, astro-ph/0106485. Cited on 111

Martín S., Martín-Pintado J., Mauersberger R. (2009). HNCO Abundances in Galaxies: Tracing the Evolutionary State of Starbursts. ApJ, 694:610–617, 0812.3688. Cited on 125

Mauersberger R., Henkel C., Weiß A., Peck A. B., Hagiwara Y. (2003). Dense gas in nearby galaxies. XV. Hot ammonia in NGC 253, Maffei 2 and IC 342. A&A, 403:561–571, astro-ph/0303247. Cited on 125

Meier D. S., Turner J. L., Hurt R. L. (2000). Molecular Gas Properties of the Starburst Nucleus of IC 342: High-Resolution $^{13}$CO (2-1) Imaging. ApJ, 531:200–209. Cited on 125

Meier D. S., Turner J. L. (2005). Spatially Resolved Chemistry in Nearby Galaxies. I. The Center of IC 342. ApJ, 618:259–280, astro-ph/0410039. Cited on 125, 126

Oberst T. E., et al. (2006). Detection of the 205 $\mu$m [N II] Line from the Carina Nebula. ApJ, 652:L125–L128, astro-ph/0610636. Cited on 128, 130

Panuzzo P., et al. (2010). Probing the molecular interstellar medium of M82 with Herschel-SPIRE spectroscopy. A&A, 518:L37, 1005.1877. Cited on 128

Papadopoulos P. P., van der Werf P., Isaak K., Xilouris E. M. (2010). CO Spectral Line Energy Distributions of Infrared-Luminous Galaxies and Active Galactic Nuclei. ApJ, 715:775–792, 1003.5889. Cited on 112

Petitpas G. R., Wilson C. D. (1998). The Physical Conditions and Dynamics of the Interstellar Medium in the Nucleus of M83: Observations of CO and C i. ApJ, 503:219, astro-ph/9803230. Cited on 112

Pilbratt G. L., et al. (2010). Herschel Space Observatory. An ESA facility for far-infrared and submillimetre astronomy. A&A, 518:L1, 1005.5331. Cited on 112

Pound M. W., Wolfire M. G. (2008). The Photo Dissociation Region Toolbox. In Argyle R. W., Bunclark P. S., Lewis J. R., editors, *Astronomical Data Analysis Software and Systems XVII*, volume 394 of *Astronomical Society of the Pacific Conference Series*, page 654. Cited on 128

Rangwala N., et al. (2011). Observations of Arp 220 Using Herschel-SPIRE: An Unprecedented View of the Molecular Gas in an Extreme Star Formation Environment. ApJ, 743:94, 1106.5054. Cited on 129, 131

Rigopoulou D., Kunze D., Lutz D., Genzel R., Moorwood A. F. M. (2002). An ISO-SWS survey of molecular hydrogen in starburst and Seyfert galaxies. A&A, 389:374–386, astro-ph/0206135. Cited on 124

Skilling J. (2004). Nested sampling. *BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEERING: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. AIP Conference Proceedings*, 735:395. Cited on 117

Solomon P. M., Vanden Bout P. A. (2005). Molecular Gas at High Redshift. ARA&A, 43:677–725, astro-ph/0508481. Cited on 130

Spinoglio L., et al. (2012). Submillimeter Line Spectrum of the Seyfert Galaxy NGC 1068 from the Herschel-SPIRE Fourier Transform Spectrometer. ApJ, 758:108, 1208.6132. Cited on 131

Stutzki J., et al. (1997). Atomic Carbon in M82: Physical Conditions Derived from Simultaneous Observations of the [C i] Fine-structure Submillimeter-wave Transitions. ApJ, 477:L33, astro-ph/9612112. Cited on 112, 129

Swinyard B. M., et al. (2010). In-flight calibration of the Herschel-SPIRE instrument. A&A, 518:L4, 1005.5073. Cited on 113

Tielens A. G. G. M., Hollenbach D. (1985). Photodissociation regions. I - Basic model. II - A model for the Orion photodissociation region. ApJ, 291:722–754. Cited on 126

van der Tak F. F. S., Black J. H., Schöier F. L., Jansen D. J., van Dishoeck E. F. (2007). A computer program for fast non-LTE analysis of interstellar line spectra. With diagnostic plots to interpret observed line intensity ratios. A&A, 468:627–635, 0704.0155. Cited on 110, 117

van der Werf P. P., et al. (2010). Black hole accretion and star formation as drivers of gas excitation and chemistry in Markarian 231. A&A, 518:L42, 1005.2877. Cited on 126, 131

Wagg J., Carilli C. L., Wilner D. J., Cox P., De Breuck C., Menten K., Riechers D. A., Walter F. (2010). [CII] line emission in BRI 1335-0417 at z = 4.4. A&A, 519:L1, 1008.1578. Cited on 130

Wilson C. D. (1997). Atomic Carbon Emission from Individual Molecular Clouds in M33. ApJ, 487:L49, astro-ph/9707158. Cited on 112, 129

Wu R., et al. (2013). Observing extended sources with the Herschel SPIRE Fourier Transform Spectrometer. A&A, 556:A116, 1306.5780. Cited on 113, 114

# Chapter 5

# A Dust-Obscured Massive Maximum-Starburst Galaxy at a Redshift of 6.34

*Dominik A. Riechers, C. M. Bradford, D. L. Clements, C. D. Dowell, I. Pérez-Fournon, R. J. Ivison, C. Bridge, A. Conley, Hai Fu, J. D. Vieira, J. Wardlow, J. Calanog, A. Cooray, P. Hurley, R. Neri, J. Kamenetzky, J. E. Aguirre, B. Altieri, V. Arumugam, D. J. Benford, M. Béthermin, J. Bock, D. Burgarella, A. Cabrera-Lavers, S. C. Chapman, P. Cox, J. S. Dunlop, L. Earle, D. Farrah, P. Ferrero, A. Franceschini, R. Gavazzi, J. Glenn, E. A. Gonzalez Solares, M. A. Gurwell, M. Halpern, E. Hatziminaoglou, A. Hyde, E. Ibar, A. Kovács, M. Krips, R. E. Lupu, P. R. Maloney, P. Martinez-Navajas, H. Matsuhara, E. J. Murphy, B. J. Naylor, H. T. Nguyen, S. J. Oliver, A. Omont, M. J. Page, G. Petitpas, N. Rangwala, I. G. Roseboom, D. Scott, A. J. Smith, J. G. Staguhn, A. Streblyanska, A. P. Thomson, I. Valtchanov, M. Viero, L. Wang, M. Zemcov & J. Zmuidzinas*

## 5.1 Introduction (from main paper)

Massive present-day early-type (elliptical and lenticular) galaxies probably gained the bulk of their stellar mass and heavy elements through intense, dust-enshrouded starbursts - that is, increased rates of star formation - in the most massive dark matter halos at early epochs. However, it remains unknown how soon after the Big Bang such massive starburst progenitors exist. The measured redshift ($z$) distribu-

tion of dusty, massive starbursts has long been suspected to be biased low in redshift owing to selection effects (Chapman et al., 2003), as confirmed by recent findings of systems out to redshift $z \sim 5$ (Capak et al., 2011; Vieira et al., 2013; Walter et al., 2012).

Here we report the identification of a massive starburst galaxy at redshift 6.34 through a submillimeter colour-selection technique. We unambiguously determined the redshift from a suite of molecular and atomic fine structure cooling lines. These measurements reveal a hundred billion solar masses of highly excited, chemically evolved interstellar medium (ISM) in this galaxy, which constitutes at least 40% of the baryonic mass. A "maximum starburst" converts the gas into stars at a rate more than 2,000 times that of the Milky Way, a rate among the highest observed at any epoch. Despite the overall downturn of cosmic star formation towards the highest redshifts (Bouwens et al., 2011), it seems that environments mature enough to form the most massive, intense starbursts existed at least as early as 880 million years after the Big Bang.

We have searched 21 deg$^2$ of the Herschel/SPIRE data of the HerMES blank field survey (Oliver et al., 2012) at 250 - 500 $\mu$m for “ultra-red” sources with flux densities $S_{250\mu\mathrm{m}} < S_{350\mu\mathrm{m}} < S_{500\mu\mathrm{m}}$ and $S_{500\mu\mathrm{m}}/S_{350\mu\mathrm{m}} > 1.3$, i.e., galaxies that are significantly redder (and thus, potentially at higher redshift) than massive starbursts discovered thus far. This selection yields five candidate ultra-red sources down to a flux limit of 30 mJy at 500 $\mu$m ($> 5\sigma$ above the confusion noise; see Supplementary Information Section 1 for additional details), corresponding to a source density of $\leq 0.24$ deg$^{-2}$. For comparison, models of number counts in the Herschel/SPIRE bands suggest a space density of massive starburst galaxies at $z > 6$ with $S_{500\mu\mathrm{m}} > 30$ mJy of 0.014 deg$^{-2}$ (Béthermin et al., 2012).

To understand the nature of galaxies selected by this technique, we have obtained full frequency scans of the 3-mm and 1-mm bands toward HFLS3 (also known as 1HERMES S350 J170647.8+584623; $S_{500\mu\mathrm{m}}/S_{350\mu\mathrm{m}} = 1.45$), the brightest candidate discovered in our study. These observations, augmented by selected follow-up over a broader wavelength range, unambiguously determine the galaxy redshift to be $z = 6.3369\pm0.0009$ based on a suite of 7 CO lines, 7 $H_2O$ lines, and OH, $OH^+$, $H_2O^+$, $NH_3$, [CI], and [CII] lines detected in emission and absorption (Figure 5.3. At this redshift, the Universe was just 880 million years old (or 1/16th of its present age),

and 1" on the sky corresponds to a physical scale of 5.6 kpc. Further observations from optical to radio wavelengths reveal strong continuum emission over virtually the entire wavelength range between 2.2 $\mu$m and 20 cm, with no detected emission short-ward of 1 $\mu$m (see Supplementary Information Section 2 and Figures S1-S11 for additional details).

My primary contribution to this paper was the radiative transfer modelling of the $H_2O$ SLED, described in Sections 5.2. I carried out all modelling of the $H_2O$ SLED with my combined setup of *Multinest* (Feroz & Hobson, 2008) and RADEX (van der Tak et al., 2007). Figure 5.2 was created by me. I also made significant contributions to both the interpretation and text in this section. My other contributions include suggestions and corrections to the remaining text throughout the paper. The following section is taken from the Supplementary information attached to Riechers et al. (2013) and is that most relevant to my contribution. Additional information on HFLS3 can be found in Section 5.3 and is taken from the main paper.

## 5.2 $H_2O$ Excitation (Section 4.2 in Supplementary information)

Given the limited constraints on the $H_2O$ excitation ladder (Figure 5.2), and the fact that the $H_2O$ line ratios in HFLS3 (within the errors) are consistent with those in Arp 220, we first explored a wider parameter space based on the $H_2O$ excitation ladder of Arp 220, and used our findings as input to models of HFLS3. We used RADEX to produce $H_2O$ radiative transfer models. The parameters and ranges explored by our models are $T_{kin} = 10^{1.3} - 10^{3.3}$ K, $n(H_2)$=$10^{3.0} - 10^{11.0}$cm$^{-3}$, $N_{H_2O} = 10^{16} - 10^{25}$cm$^{-2}$, $\Phi_A = 10^{-5} - 1$, and [$H_2O$/$H_2$]=$10^{-9} - 10^{-5}$. As priors, we assume that the predicted molecular gas mass cannot exceed the dynamical mass, and that the $H_2O$ column length is smaller than or equal to the observed diameter of the CO and $H_2O$-emitting region. We also fixed the $H_2O$ ortho-to-para ratio to 3, and reject RADEX runs where the optical depth of all lines is outside the range $\tau = 10^{-10} - 10^{2}$. We generated posterior likelihood distributions for all model parameters using the nested sampling routine MULTINEST, which we used to call RADEX and create models "on-the-fly" (Feroz & Hobson, 2008). MULTINEST is a Bayesian tool similar to MCMC, but is more efficient when dealing with multi-modal regions in parameter

space. It also calculates values for the Bayesian evidence. For both Arp 220 and HFLS3, [$H_2O/H_2$], and thus, $M_{gas}$, are poorly constrained by the $H_2O$ data alone. Solutions with $T_{kin}$ ∼100-150 K are preferred, comparable to the best-fitting values for CO in HFLS3. However, due to the high critical densities of the $H_2O$ lines, gas densities of $>10^{8.5} cm^{-3}$ are preferred (black contours in Figure 5.2). Such high gas densities over extended regions are unphysical. Also, the maximum likelihood solutions only poorly reproduce the observed line ratios in Arp 220 and HFLS3, in particular between directly connected energy levels (as higher energy levels are observed to have higher line fluxes in both Arp 220 and HFLS3; Figure 5.2). Our models thus suggest that collisional excitation is unlikely to be the main mechanism to produce the observed $H_2O$ excitation in Arp 220 and HFLS3.

To further explore whether or not collisional excitation is important for lower energy levels, we carried out a separate modeling run, taking only fluxes and limits for lines with $E_{upper}/k_B$ <∼200 K into account. Solutions with less extreme gas densities of $10^7 - 10^8 \mathrm{cm}^{-3}$ are preferred (red contours in Figure S16), but at significantly higher $T_{kin}$ of few hundreds to >1000 K. The best-fitting models still only poorly reproduce the observed line ratios of directly connected energy levels. Thus, it appears unlikely that collisional excitation dominates the lower-energy $H_2O$ levels either.

We detect emission from $H_2O$ lines with upper level energies of $E/k_B > 300$ - 450 K, which would require gas densities in excess of $10^{8.5} \mathrm{cm}^{-3}$ to reproduce the observed relative line strengths through collisions. However, the $J_{KaKc} = 3_{21}$ and $4_{22}$ energy levels of ortho- and para-$H_2O$ can be efficiently populated by 75 and 58 $\mu$m infrared photons through absorption in the $J_{KaKc} = 2_{12} - 3_{21}$ and $3_{13} - 4_{22}$ ortho- and para-$H_2O$ transitions, respectively, which then can produce the observed emission line strength in the cascading transitions. The 75 $\mu$m transition coincides with the observed peak of the SED, allowing for a high pumping efficiency. The 58 $\mu$m transition falls on the Wien tail of the SED. In combination with the higher energy of the lower level of the pumping transition (and thus, lower population through collisions) compared to the ortho-$H_2O$ channel, this explains the much lower strength of high-level para-$H_2O$ lines. The relative strength of all detected $H_2O$ lines is consistent with those observed in Arp 220 (Fig. 3) (Rangwala et al., 2011). The 75 $\mu$m and 58 $\mu$m pumping transitions were detected in absorption in Arp 220, with

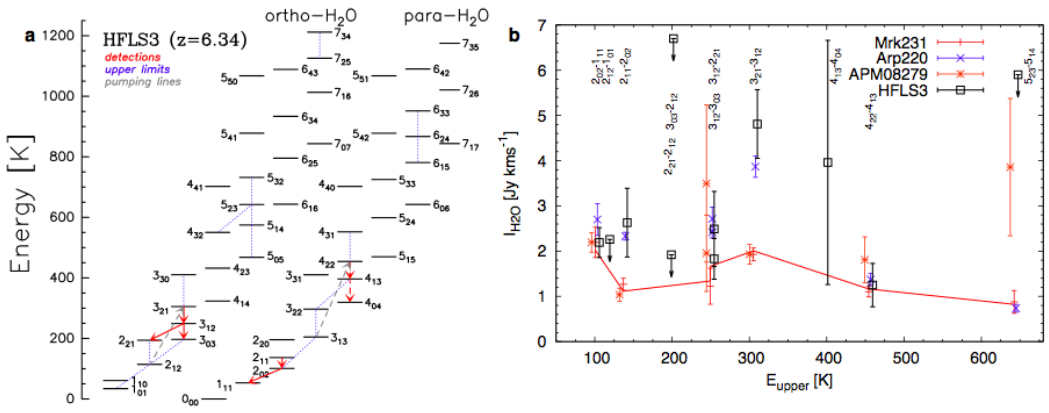


Figure 5.1: Energy level diagram **(a)** and observed line intensities textbf(b) for $H_2O$ in HFLS3. **a**, The red arrows indicate detected $H_2O$ transitions (dashed lines indicate tentative detections) and blue dotted lines indicate upper limits. The dashed gray arrows indicate 75 and 58 $\mu$m infrared pumping transitions of ortho- and para-$H_2O$ that can efficiently radiatively populate the energy levels from which bright emission lines are observed. **b**, The black squares indicate observed $H_2O$ line fluxes and upper limits (shallow limits not shown) for HFLS3 as a function of upper level energy. For comparison the red, blue and orange symbols show normalised line fluxes observed in Mrk 231 (red line: model), Arp 220, and APM08279+5255 (Rangwala et al., 2011; Bradford et al., 2011; González-Alfonso et al., 2012; Thompson et al., 2005). Within the uncertainties, the $H_2O$ excitation in HFLS3 is comparable to that in Arp 220, but inconsistent with the line ratios in the quasars Mrk 231 and APM08279+5255 (where the $H_2O$ lines may be substantially excited by a strong X-ray radiation field associated with their AGN). In particular, the relative strength of the $J_{KaKc}$=$3_{21} - 3_{12}$ and $2_{11} - 2_{02}$ lines relative to transitions lower in the cascade in HFLS3 suggest that their upper energy levels may be substantially populated through radiative excitation by the infrared radiation field in the star-forming regions, rather than collisions. Given the median gas density of $\sim 10^4$ cm$^{-3}$ as determined from the CO excitation and the critical densities of $> 10^9$ cm$^{-3}$ for the detected $H_2O$ lines, contributions from collisional excitation to the observed line fluxes are likely minor. This suggests that the $H_2O$ emission lines detected in HFLS3 are less important for gas cooling through the removal of kinetic energy than the CO lines.

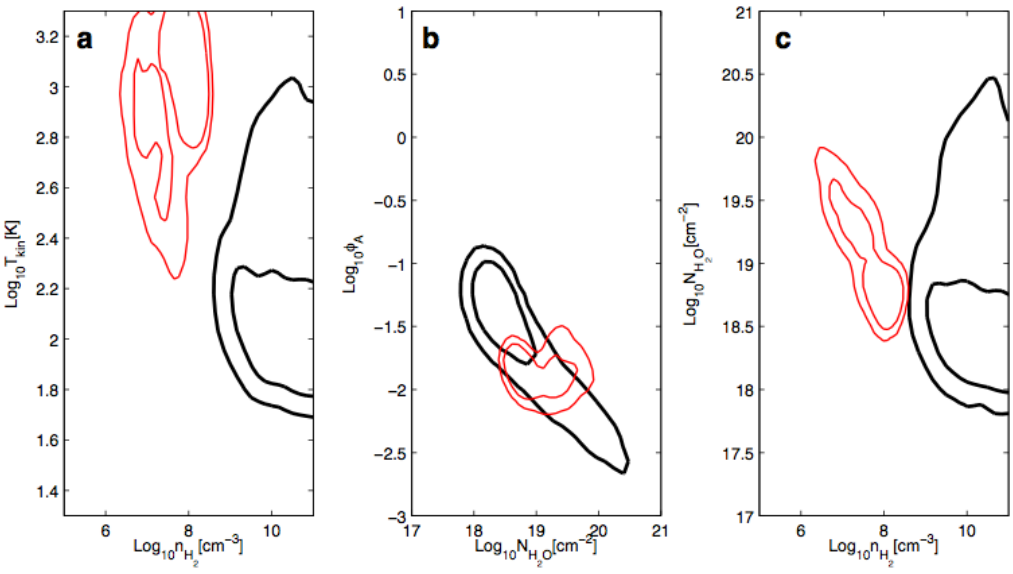


Figure 5.2: Physical parameters obtained from fitting the $H_2O$ excitation ladder of HFLS3. Two-dimensional likelihood contours for the physical properties of the molecular gas in HFLS3 as determined by RADEX models of the $H_2O$ excitation ladder for $n(H_2)$ and $T_{kin}$ **(a)**, $N_{H_2O}$ and $\Phi_A$ **(b)**, and $n(H_2)$ and $N_{H_2O}$ **(c)**. Black contours show the likelihoods when fitting all observed $H_2O$ lines and limits. Red contours show the likelihoods when only considering lines with $E_{upper}/k_B <\sim 200$K. Contour levels are shown at 1 and $2\sigma$.

the 75 $\mu$m feature being more than twice as deep (González-Alfonso et al., 2012). This is also consistent with the fact that we detect the OH $^2\Pi_{1/2}$(3/2-1/2) feature at 163 $\mu$m in emission, which has an upper level energy comparable to the $J_{KaKc} = 3_{21}$ level of ortho-$H_2O$, but can also be efficiently pumped by 53.3 and 35 $\mu$m photons, as observed in Arp 220 (González-Alfonso et al., 2012).

Our models thus suggest that collisional excitation is unlikely to be the main mechanism to produce the observed $H_2O$ excitation in HFLS3, which instead may be enhanced by the infrared radiation field in the star-forming regions. This is consistent with observations and models of very high energy levels of $H_2O$ in Arp 220, which are observed in absorption (González-Alfonso et al., 2012).

## 5.3 Additional Information on HFLS3 (from the main paper)

HFLS3 hosts an intense starburst. The 870 $\mu$m-flux of HFLS3 is $> 3.5$ times higher than those of the brightest high-redshift starbursts in a 0.25-deg$^2$ region containing the Hubble Ultra Deep Field (HUDF) (Karim et al., 2013). From the continuum spectral energy distribution (Figure 5.4), we find that the far-infrared (FIR) luminosity $L_{FIR}$ and inferred star formation rate (SFR) of 2,900 $M_\odot yr^{-1}$ of HFLS3 are 15-20 times those of the prototypical local ultra-luminous starburst Arp 220, and $>$ 2,000 times those of the Milky Way (Table 5.1 and Supplementary Information Section 3). The SFR of HFLS3 alone corresponds to $\sim 4.5$ times the ultraviolet-based SFR of all $z = 5.5 - 6.5$ star-forming galaxies in the HUDF combined (Bouwens et al., 2006), but the rarity and dust obscuration of ultra-red sources like HFLS3 implies that they do not dominate the UV photon density needed to reionize the Universe (Robertson et al., 2010).

HFLS3 is a massive, gas-rich galaxy. From the spectral energy distribution and the intensity of the CO and [CII] emission, we find a dust mass of $M_d = 1.3 \times 10^9 M_\odot$ and total molecular and atomic gas masses of $M_{gas} = 1.0 \times 10^{11} M_\odot$ and $M_{HI} = 2.0 \times 10^{10} M_\odot$.

These masses are 15-20 times those of Arp 220, and correspond to a gas-to-dust ratio of $\sim 80$ and a gas depletion timescale of $M_{gas}$/SFR $\sim$ 36Myr. These values are comparable to lower-redshift submillimeter-selected starbursts (Michałowski et al.,

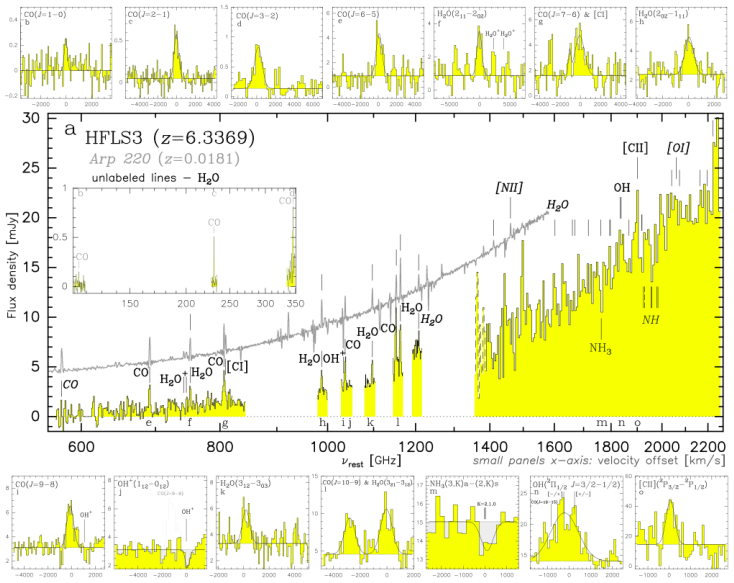


Figure 5.3: Redshift identification through molecular and atomic spectroscopy of HFLS3. **a**, Black trace, wide-band spectroscopy in the observed-frame 19 - 0.95-mm (histogram; rest-frame 2,600 - 130 $\mu$m) wavelength range with CARMA (3 mm; blind frequency scan of the full band), the PdBI (2 mm), the JVLA (19 - 6 mm), and CSO/Z-spec (1 mm; instantaneous coverage). (CARMA, Combined Array for Research in Millimeter-wave Astronomy; PdBI, Plateau de Bure Interferometer; JVLA, Jansky Very Large Array; and CSO, Caltech Submillimeter Observatory) This uniquely determines the redshift of HFLS3 to be $z$=6.3369 based on the detection of a series of $H_2O$, CO, OH, $OH^+$, NH3, [CI] and [CII] emission and absorption lines. **b** to **o**, Detailed profiles of detected lines (histograms; rest frequencies are indicated by corresponding letters in **a)**. 1-mm lines **(m-o)** are deeper, interferometric confirmation observations for NH3, OH (both PdBI), and [CII] (CARMA) not shown in **a**. The line profiles are typically asymmetric relative to single Gaussian fits, indicating the presence of two principal velocity components at redshifts of 6.3335 and 6.3427. The implied CO, [CI], and [CII] line luminosities are $5.08\pm0.45\times10^6$, $3.0\pm1.9\times10^8$, and $1.55\pm0.32\times10^{10}L_\odot$. Strong rest-frame submillimeter to far-infrared continuum emission is detected over virtually the entire wavelength range. For comparison, the Herschel/SPIRE spectrum of the nearby ultra-luminous infrared galaxy Arp 220 (Rangwala et al., 2011) is overplotted in grey **(a)**. Lines labeled in italic are tentative detections or upper limits (see Table S2). Most of the bright spectral features detected in Arp 220 (González-Alfonso et al., 2012; Rangwala et al., 2011) are also detected in HFLS3 (in spectral regions not blocked by the terrestrial atmosphere). See Supplementary Information Sections 2-4 for more details.

Table 5.1: Observed and derived quantities for HFLS 3, Arp 220 and the Milky Way

| | HFLS3 | Arp 220* | Milky Way* |
|---|---|---|---|
| $z$ | 6.3369 | 0.0181 | |
| $M_{gas}(M_\odot)$ † | $(1.04 \pm 0.09) \times 10^{11}$ | $5.2 \times 10^9$ | $2.5 \times 10^9$ |
| $M_{dust}(M_\odot)$‡ | $1.31^{+0.32}_{-0.30} \times 10^9$ | $\sim 1 \times 10^8$ | $\sim 6 \times 10^7$ |
| $M_*(M_\odot)$§ | $\sim 3.7 \times 10^{10}$ | $\sim (3-5) \times 10^{10}$ | $\sim 6.4 \times 10^{10}$ |
| $M_{dyn}(M_\odot)$\|\| | $2.7 \times 10^{11}$ | $3.45 \times 10^{10}$ | $2 \times 10^{11}(< 20\text{kpc})$ |
| $f_{gas}(\%)$¶ | 40 | 15 | 1.2 |
| $L_{FIR}(L_\odot)$ # | $2.86^{+0.32}_{-0.31} \times 10^{13}$ | $1.8 \times 10^{12}$ | $1.1 \times 10^{10}$ |
| SFR($M_\odot \text{yr}^{-1}$)⊛ | 2900 | $\sim 180$ | 1.3 |
| $T_{dust}$(K)** | $55.9^{+9.3}_{-12.0}$ | 66 | $\sim 19$ |

*Literature values for Arp 220 and the Milky Way are adopted from Downes & Solomon (1998); McMillan (2011); Murray & Rahman (2010); Sodroski et al. (1994). The total molecular gas mass of the Milky Way is uncertain by at least a factor of 2. Quoted dust masses and stellar masses are typically uncertain by factors of 2-3 due to systematics. The dynamical mass for the Milky Way is quoted within the inner 20 kpc to be comparable to the other systems, not probing the outer regions dominated by dark matter. The dust temperature in the Milky Way varies by at least ±5 K around the quoted value, which is used as a representative value. Both Arp 220 and the Milky Way are known to contain small fractions of significantly warmer dust. All error bars are $1\sigma$ r.m.s. uncertainties.

†Molecular gas mass, derived assuming $\alpha\text{CO} = M_{gas}/LCO = 1M_\odot(\text{K kms}^{-1}\text{pc}^2)^{-1}$, see Supplementary Information, Section 3.3.

‡Dust mass, derived from spectral energy distribution fitting (see Supplementary Information section 3.1).

§Stellar mass, derived from population synthesis fitting (see Supplementary Information section 3.4).

||Dynamical mass (see Supplementary Information section 3.5).

¶Gas mass fraction, derived assuming $f_{gas} = M_{gas}/M_{dyn}$ (see Supplementary Information section 3.6)

# FIR luminosity as determined over the range of 42.5 - 122.5$\mu$m from spectral energy distribution fitting (see Supplementary Information section 3.1).

⊛ SFR, derived assuming SFR (in $M_\odot \text{yr}^{-1}$) $=1.0\times10^{-10}L_{FIR}$ (in $L_\odot$) (see Supplementary Information section 3.2).

** Dust temperature, derived from spectral energy distribution fitting (see Supplementary Information section 3.1).

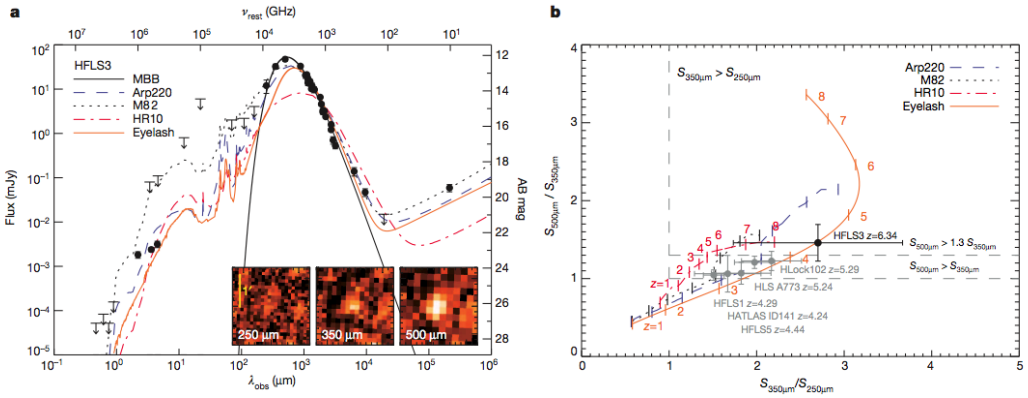


Figure 5.4: Spectral energy distribution (SED) and Herschel/SPIRE colours of HFLS3. **a**, HFLS3 was identified as a very high redshift candidate, as it appears red between the Herschel/SPIRE 250-, 350-, and 500-$\mu$m bands (inset). The SED of the source (data points; $\lambda_{obs}$, observed-frame wavelength; $\nu_{rest}$, rest-frame frequency; AB mag, magnitudes in the AB system; error bars are $1\sigma$ r.m.s. uncertainties in both panels) is fitted with a modified black body (MBB; solid line) and spectral templates for the starburst galaxies Arp 220, M82, HR10, and the Eyelash (broken lines, see key). The implied FIR luminosity is $2.86^{+0.32}_{-0.31} \times 10^{13} L_{\odot}$. The dust in HFLS3 is not optically thick at wavelengths longward of rest-frame 162.7 $\mu$m (95.4% confidence; Figure S12). This is in contrast to Arp 220, in which the dust becomes optically thick (i.e., $\tau_d = 1$) shortward of $234 \pm 3\mu$m (Rangwala et al., 2011). Other high-redshift massive starburst galaxies (including the Eyelash) typically become optically thick around $\sim$200 $\mu$m. This suggests that none of the detected molecular/fine structure emission lines in HFLS3 require correction for extinction. The radio continuum luminosity of HFLS3 is consistent with the radio-FIR correlation for nearby star-forming galaxies. **b**, 350 $\mu$m/250 $\mu$m and 500 $\mu$m/350 $\mu$m flux density ratios of HFLS3. The coloured lines are the same templates as in **a**, but redshifted between $1 < z < 8$ (number labels indicate redshifts). Dashed grey lines indicate the dividing lines for red ($S_{250\mu m} < S_{350\mu m} < S_{500\mu m}$) and ultra-red sources ($S_{250\mu m} < S_{350\mu m}$ and $1.3 \times S_{350\mu m} < S_{500\mu m}$). Gray symbols show the positions of five spectroscopically confirmed red sources at $4 < z < 5.5$ (including three new sources from our study), which all fall outside the ultra-red cutoff. This shows that ultra-red sources will lie at $z > 6$ for typical SED shapes (except those with low dust temperatures), while red sources typically are at $z < 5.5$. See Supplementary Information Sections 1 and 3 for more details.

2010; Riechers et al., 2011). From the [CI] luminosity, we find an atomic carbon mass of $4.5 \times 10^7 M_\odot$. At the current star formation rate of HFLS3, this level of carbon enrichment could have been achieved through supernovae on a timescale of $\sim 10^7$yr (Walter et al., 2003). The profiles of the molecular and atomic emission lines typically show two velocity components (Figs. 5.3, S5, and S7). The gas is distributed over a 1.7 kpc radius region with a high velocity gradient and dispersion (Figure 5.5). This suggests a dispersion-dominated galaxy with a dynamical mass of $M_{dyn} = 2.7 \times 10^{11} M_\odot$. The gas mass fraction in galaxies is a measure of the relative depletion and replenishment of molecular gas, and is expected to be a function of halo mass and redshift from simulations (Lagos et al., 2011). In HFLS3, we find a high gas mass fraction of $f_{gas} = M_{gas}/M_{dyn} \sim 40\%$, comparable to what is found in submillimeter-selected starbursts and massive star-forming galaxies at $z \sim$2,(Tacconi et al., 2008, 2010) but $\sim$3 times higher than in nearby ultra-luminous infrared galaxies (ULIRGs) like Arp 220, and >30 times higher than in the Milky Way. From population synthesis modeling, we find a stellar mass of $M_* = 3.7 \times 10^{10} M_\odot$ , comparable to that of Arp 220 and about half that of the Milky Way. This suggests that at most $\sim$40% of $M_{dyn}$ within the radius of the gas reservoir are due to dark matter. With up to $\sim 10^{11} M_\odot$ of dark matter within 3.4 kpc, HFLS3 likely resides in a dark matter halo massive enough to grow a present-day galaxy cluster (Overzier et al., 2009).

The efficiency for star formation is given by $\epsilon = t_{dyn} \times \mathrm{SFR}/M_{gas}$, where $t_{dyn} = r^3/(2GM))^{1/2}$ is the dynamical (or free-fall) time, $r$ is the source radius, $M$ is the mass within radius $r$ and $G$ is the gravitational constant. For $r$=1.7kpc and $M = M_{gas}$, this suggests $\epsilon$=0.06, which is a few times higher than found in nearby starbursts and in Giant Molecular Cloud cores in the Galaxy (Krumholz et al., 2012).

The properties of the atomic and molecular gas in HFLS3 are fully consistent with a highly enriched, highly excited interstellar medium, as typically found in the nuclei of warm, intense starbursts, but distributed over a large, $\sim$3.5-kpc-diameter region. The observed CO and [CII] luminosities suggest that dust is the primary coolant of the gas if both are thermally coupled. The $L_{[CII]}/L_{FIR}$ ratio of $\sim 5 \times 10^{-4}$ is typical for high radiation environments in extreme starbursts and active galactic nucleus (AGN) host galaxies (Stacey et al., 2010). The $L_{[CII]}/L_{CO(1-0)}$ ratio of

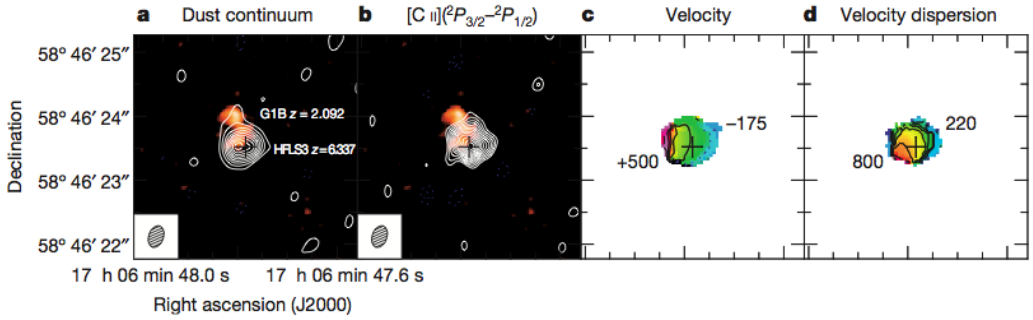


Figure 5.5: Gas dynamics, dust obscuration, and distribution of gas and star formation in HFLS3. **a**, **b**, High-resolution (FWHM 0.35"×0.2") maps of the 158-$\mu$m continuum **(a)** and [CII] line emission **(b)** obtained at 1.16 mm with the PdBI in A-configuration, overlaid on a Keck/NIRC2 2.2-$\mu$m adaptive optics image (rest-frame UV/optical light). The r.m.s. uncertainty in the continuum **(a)** and line **(b)** maps is 180 and 400 $\mu$Jy beam$^{-1}$, and contours are shown in steps of 3 and 1$\sigma$, starting at 5 and 3$\sigma$, respectively. A $z = 2.092$ galaxy (labeled G1B) identified through Keck/LRIS spectroscopy is detected ∼0.65" north of HFLS3, but is not massive enough to cause significant gravitational lensing at the position of HFLS3. Faint infrared emission is detected toward a region with lower dust obscuration in the north-eastern part of HFLS3 (not detected at $< 1\mu$m). The Gaussian diameters of the resolved [CII] and continuum emission are 3.4 kpc × 2.9 kpc and 2.6 kpc × 2.4 kpc, suggesting gas and SFR surface densities of $\Sigma_{gas} = 1.4 \times 10^4 M_{\odot} pc^{-2}$ and $\Sigma_{SFR} = 600 M_{\odot}$ yr$^{-1}$ kpc$^{-2}$ ($\sim 0.6 \times 10^{13} L_{\odot} kpc^{-2}$). The high $\Sigma_{SFR}$ is consistent with a maximum starburst at near-Eddington-limited intensity. Given the moderate optical depth of $\tau_d <\sim 1$ at 158 $\mu$m, this estimate is somewhat conservative. Peak velocity **(c)** and F.W.H.M. velocity dispersion **(d)** maps of the [CII] emission are obtained by Gaussian fitting to the line emission in each spatial point of the map. Velocity contours are shown in steps of 100 kms$^{-1}$. High-resolution CO J=7-6 and 10-9 and $H_2O$ $3_{21} - 3_{12}$ observations show consistent velocity profiles and velocity structure (Figures S5-S7). The large velocity dispersion suggests that the gas dynamics in this system are dispersion-dominated. See Supplementary Information Sections 3 and 5 for more details.

$\sim$3,000 suggests that the bulk of the line emission is associated with the photon dominated regions of a massive starburst. At the $L_{FIR}$ of HFLS3, this suggests an infrared radiation field strength and gas density comparable to nearby ULIRGs without luminous AGN (Figs. 4 & 5 of Stacey et al. (2010)).

From the spectral energy distribution of HFLS3, we derive a dust temperature of $T_{dust} = 56^{+9}_{-12}$K, $\sim$10 K less than in Arp 220, but $\sim$3 times that of the Milky Way. CO radiative transfer models assuming collisional excitation suggest a gas kinetic temperature of $T_{kin} = 144^{+59}_{-30}$K and a gas density of $\log_{10}(\mathrm{n(H_2)})=3.80^{+0.28}_{-0.17}$ cm$^{-3}$ (Supplementary Information Section 4 and Figures S13/S14). These models suggest similar gas densities as in nearby ULIRGs, and prefer $T_{kin} \gg T_{dust}$, which may imply that the gas and dust are not in thermal equilibrium, and that the excitation of the molecular lines may be partially supported by the underlying infrared radiation field. This is consistent with the finding that we detect $H_2O$ and OH lines with upper level energies of $E/k_B >$ 300-450 K and critical densities of $>10^{8.5}$cm$^{-3}$ at line intensities exceeding those of the CO lines. The intensities and ratios of the detected $H_2O$ lines cannot be reproduced by radiative transfer models assuming collisional excitation, but are consistent with being radiatively pumped by far-infrared photons, at levels comparable to those observed in Arp 220 (Figures S15/S16) (González-Alfonso et al., 2012; Rangwala et al., 2011). The CO and $H_2O$ excitation is inconsistent with what is observed in quasar host galaxies like Mrk 231 and APM08279+5255 at $z$=3.9, which lends support to the conclusion that the gas is excited by a mix of collisions and infrared photons associated with a massive, intense starburst, rather than hard radiation associated with a luminous AGN (Thompson et al., 2005). The physical conditions in the ISM of HFLS3 thus are comparable to those in the nuclei of the most extreme nearby starbursts, consistent with the finding that it follows the radio-FIR correlation for star-forming galaxies.

HFLS3 is rapidly assembling its stellar bulge through star formation at surface densities close to the theoretically predicted limit for "maximum starbursts" (Elmegreen, 1999). At a rest-frame wavelength of 158 $\mu$m, the FIR emission is distributed over a relatively compact area of 2.6 kpc$\times$2.4 kpc physical diameter along its major and minor axes respectively (Fig. 3; as determined by elliptical Gaussian fitting).

This suggests an extreme star formation rate surface density of $\Sigma_{SFR} \sim 600 M_\odot$

$yr^{-1}kpc^{-2}$ over a 1.3-kpc-radius region, and is consistent with near-Eddington-limited star formation if the starburst disk is supported by radiation pressure (Thompson et al., 2005). This suggests the presence of a kiloparsec-scale hyper-starburst similar to that found in the z=6.42 quasar J1148+5251 (Walter et al., 2009). Such high $\Sigma_{SFR}$ are also observed in the nuclei of local ULIRGs such as Arp 220, albeit on two orders of magnitude smaller scales. A starburst at such high $\Sigma_{SFR}$ may produce strong winds. Indeed, the relative strength and broad, asymmetric profile of the OH $^2\Pi_{1/2}$(3/2-1/2) doublet detected in HFLS3 may indicate a molecular outflow, reminiscent of the OH outflow in Arp 220 (González-Alfonso et al., 2012).

The identification of HFLS3 alone is still consistent with the model-predicted space density of massive starburst galaxies at $z$ >6 with $S_{500\mu m} > 30$ mJy of 0.014 $deg^{-2}$ (Béthermin et al., 2012). This corresponds to only $10^{-3} - 10^{-4}$ times the space density of Lyman-break galaxies at the same redshift, but is comparable to the space density of the most luminous quasars hosting supermassive black holes (SMBHs, i.e., a different population of massive galaxies) at such early cosmic times (Jiang et al., 2009). The host galaxies around these very distant SMBHs are commonly FIR-luminous, but less intensely star-forming, with typically a few times lower $L_{FIR}$ than ultra-red sources (Walter et al., 2009). This highlights the difference between selecting massive z>6 galaxies at the peak of their star formation activity through $L_{FIR}$, and at the peak of their black hole activity through luminous AGN. The substantial population of ultra-red sources discovered with Herschel will be an ideal probe of early galaxy evolution and heavy element enrichment within the first billion years of cosmic time. These galaxies are unlikely to dominate the star formation history of the Universe at z>6, (Béthermin et al., 2012) but they trace the highest peaks in SFR at early epochs. A detailed study of this galaxy population will reveal the mass and redshift distribution, number density and likely environments of such objects, which if confirmed in larger numbers may present a stern challenge to current models of early cosmic structure formation.

## Acknowledgments

Herschel is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA. This research has made use of data from HerMES project (`http://hermes.sussex.`

ac.uk/). HerMES is a Herschel Key Programme utilising Guaranteed Time from the SPIRE instrument team, ESAC scientists and a mission scientist. See Supplementary Information for further acknowledgments.

**Author Contributions**

D.R. had the overall lead of the project. C.M.B., D.C., I.P.-F., R.I., C.B., H.F., J.V., and R.N. have contributed significantly to the taking and analysis of the follow-up data with different instruments by leading several telescope proposals and analysis efforts. C.D.D. has led the selection of the parent sample. A.C., J.W., J.C., A.C., P.H., and J.K. have contributed significantly to the data analysis and to fitting and modeling the results. All other authors contributed to the proposals, source selection, data analysis and interpretation, in particular through work on the primary Herschel SPIRE data in which the source was discovered through the HerMES consortium (led by J.B. and S.O.). All authors have reviewed, discussed, and commented on the manuscript.

# Bibliography

Béthermin M., et al. (2012). A Unified Empirical Model for Infrared Galaxy Counts Based on the Observed Physical Evolution of Distant Galaxies. ApJ, 757:L23, 1208.6512. Cited on 139, 151

Bouwens R. J., Illingworth G. D., Blakeslee J. P., Franx M. (2006). Galaxies at z ˜ 6: The UV Luminosity Function and Luminosity Density from 506 HUDF, HUDF Parallel ACS Field, and GOODS i-Dropouts. ApJ, 653:53–85, astro-ph/0509641. Cited on 144

Bouwens R. J., et al. (2011). A candidate redshift z˜10 galaxy and rapid changes in that population at an age of 500Myr. Nature, 469:504–507, 0912.4263. Cited on 139

Bradford C. M., et al. (2011). The Water Vapor Spectrum of APM 08279+5255: X-Ray Heating and Infrared Pumping over Hundreds of Parsecs. ApJ, 741:L37, 1106.4301. Cited on 142

Capak P. L., et al. (2011). A massive protocluster of galaxies at a redshift of z˜5.3. Nature, 470:233–235, 1101.3586. Cited on 139

Chapman S. C., Blain A. W., Ivison R. J., Smail I. R. (2003). A median redshift of 2.4 for galaxies bright at submillimetre wavelengths. Nature, 422:695–698, astro-ph/0304235. Cited on 139

Downes D., Solomon P. M. (1998). Rotating Nuclear Rings and Extreme Starbursts in Ultraluminous Galaxies. ApJ, 507:615–654, astro-ph/9806377. Cited on 146

Elmegreen B. G. (1999). Galactic Bulge Formation as a Maximum Intensity Starburst. ApJ, 517:103–107, astro-ph/9901025. Cited on 150

Feroz F., Hobson M. P. (2008). Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. MNRAS, 384:449–463, 0704.3704. Cited on 140

González-Alfonso E., et al. (2012). Herschel/PACS spectroscopy of NGC 4418 and Arp 220: $H_2O$, $H_2{}^{18}O$, OH, $^{18}OH$, O I, HCN, and $NH_3$. A&A, 541:A4, 1109.1118. Cited on 142, 144, 145, 150, 151

Jiang L., et al. (2009). A Survey of z $\sim$ 6 Quasars in the Sloan Digital Sky Survey Deep Stripe. II. Discovery of Six Quasars at z $_{AB}$ >21. AJ, 138:305–311, 0905.4126. Cited on 151

Karim A., et al. (2013). An ALMA survey of submillimetre galaxies in the Extended Chandra Deep Field South: high-resolution 870 $\mu$m source counts. MNRAS, 432:2–9, 1210.0249. Cited on 144

Krumholz M. R., Dekel A., McKee C. F. (2012). A Universal, Local Star Formation Law in Galactic Clouds, nearby Galaxies, High-redshift Disks, and Starbursts. ApJ, 745:69, 1109.4150. Cited on 148

Lagos C. D. P., Lacey C. G., Baugh C. M., Bower R. G., Benson A. J. (2011). On the impact of empirical and theoretical star formation laws on galaxy formation. MNRAS, 416:1566–1584, 1011.5506. Cited on 148

McMillan P. J. (2011). Mass models of the Milky Way. MNRAS, 414:2446–2457, 1102.4340. Cited on 146

Michałowski M. J., Watson D., Hjorth J. (2010). Rapid Dust Production in Submillimeter Galaxies at z $<$ 4? ApJ, 712:942–950, 1002.2636. Cited on 144

Murray N., Rahman M. (2010). Star Formation in Massive Clusters Via the Wilkinson Microwave Anisotropy Probe and the Spitzer Glimpse Survey. ApJ, 709:424–435, 0906.1026. Cited on 146

Oliver S. J., et al. (2012). The Herschel Multi-tiered Extragalactic Survey: HerMES. MNRAS, 424:1614–1635, 1203.2562. Cited on 139

Overzier R. A., Guo Q., Kauffmann G., De Lucia G., Bouwens R., Lemson G. (2009). $\Lambda$CDM predictions for galaxy protoclusters - I. The relation between

galaxies, protoclusters and quasars at z $\sim$ 6. MNRAS, 394:577–594, 0810.2566. Cited on 148

Rangwala N., et al. (2011). Observations of Arp 220 Using Herschel-SPIRE: An Unprecedented View of the Molecular Gas in an Extreme Star Formation Environment. ApJ, 743:94, 1106.5054. Cited on 141, 142, 145, 147, 150

Riechers D. A., Hodge J., Walter F., Carilli C. L., Bertoldi F. (2011). Extended Cold Molecular Gas Reservoirs in z $\sim=$ 3.4 Submillimeter Galaxies. ApJ, 739:L31, 1105.4177. Cited on 148

Riechers D. A., et al. (2013). A dust-obscured massive maximum-starburst galaxy at a redshift of 6.34. Nature, 496:329–333, 1304.4256. Cited on 140

Robertson B. E., Ellis R. S., Dunlop J. S., McLure R. J., Stark D. P. (2010). Early star-forming galaxies and the reionization of the Universe. Nature, 468:49–55, 1011.0727. Cited on 144

Sodroski T. J., et al. (1994). Large-scale characteristics of interstellar dust from COBE DIRBE observations. ApJ, 428:638–646. Cited on 146

Stacey G. J., Hailey-Dunsheath S., Ferkinhoff C., Nikola T., Parshley S. C., Benford D. J., Staguhn J. G., Fiolet N. (2010). A 158 $\mu$m [C II] Line Survey of Galaxies at z $\sim$ 1-2: An Indicator of Star Formation in the Early Universe. ApJ, 724:957–974, 1009.4216. Cited on 148, 150

Tacconi L. J., et al. (2008). Submillimeter Galaxies at z ˜ 2: Evidence for Major Mergers and Constraints on Lifetimes, IMF, and CO-$H_2$ Conversion Factor. ApJ, 680:246–262, 0801.3650. Cited on 148

Tacconi L. J., et al. (2010). High molecular gas fractions in normal massive star-forming galaxies in the young Universe. Nature, 463:781–784, 1002.2149. Cited on 148

Thompson T. A., Quataert E., Murray N. (2005). Radiation Pressure-supported Starburst Disks and Active Galactic Nucleus Fueling. ApJ, 630:167–185, astro-ph/0503027. Cited on 142, 150, 151

van der Tak F. F. S., Black J. H., Schöier F. L., Jansen D. J., van Dishoeck E. F. (2007). A computer program for fast non-lte analysis of interstellar line spectra.

with diagnostic plots to interpret observed line intensity ratios. *Astronomy and Astrophysics*, 468:627. Cited on 140

Vieira J. D., et al. (2013). Dusty starburst galaxies in the early Universe as revealed by gravitational lensing. Nature, 495:344–347, 1303.2723. Cited on 139

Walter F., et al. (2003). Molecular gas in the host galaxy of a quasar at redshift z = 6.42. Nature, 424:406–408, astro-ph/0307410. Cited on 148

Walter F., Riechers D., Cox P., Neri R., Carilli C., Bertoldi F., Weiss A., Maiolino R. (2009). A kiloparsec-scale hyper-starburst in a quasar host less than 1gigayear after the Big Bang. Nature, 457:699–701, 0902.0662. Cited on 151

Walter F., et al. (2012). The intense starburst HDF850.1 in a galaxy overdensity at z=5.2 in the Hubble Deep Field. Nature, 486:233–236, 1206.2641. Cited on 139

# Chapter 6

# Discussion and Conclusions

The papers contained within this thesis have used two approaches to help untangle the contributions made from star formation and AGN: multivariate analysis and theoretical modelling. In this concluding chapter I will summarise what has been learnt from both, the advantages and pitfalls and the potential for future work.

## 6.1 Multivariate analysis

In the first two papers of this thesis, I have shown how multivariate analysis techniques can be successfully used to analyse the mid-infrared spectra of galaxies.

In the first paper, I showed that most of the information contained within the IRS spectra of 119 local ULIRGs could be explained by five principal components. This leads to a practical advantage in terms of reducing the dimensionality of ULIRG spectra to five numbers, enabling other techniques such as Gaussian mixtures modelling to be used for defining classification schemes. Although this can be seen as a more robust approach as the entire MIR region is used rather than only a few spectral features (e.g. Spoon et al., 2007), the advantages are mostly practical and statistical in nature. What have we learnt in terms of astrophysics?

The first key point is that with an objective approach, rather than with a subjective prejudice, I have shown the ULIRG spectra can be explained by five principal components, implying the natural dimensionality of ULIRG spectra will be of a similar order of magnitude. The next obvious question is what physical processes drive this natural dimensionality? Examining what features are present in each principal component can give some indication of what each PC does. For example the PAH features in PC2 and PC3 would suggest that both PCs are related to star formation.

I also showed how the physical parameters of radiative transfer models were related to some of the components, for example the viewing angle of the dusty AGN torus to PC1 and PC4, AGN/star formation power ratio to PC3. Despite some success, the interpretation still remains unclear. The presence of the same features in multiple components, and the non-linear behaviour of PC contribution with physical parameters (e.g. PC4 and viewing angle in Figure 2.10) suggests that the assumptions required by PCA are not well suited for continuos mid-infrared spectral data.

Paper two showed other techniques were more suited for mid-infrared spectra of galaxies. Section 3.4 discussed simple models for mid-infrared spectra, and how NMF was more appropriate than PCA or the other popular technique, independent component analysis. This paper was also more ambitious in that spectra of all extragalactic sources within the redshift range ($0.01 < z < 0.2$) were used as the input. By using a more appropriate technique, the derived components are more physically intuitive.

Now we have the added advantage of reducing the dimensionality of galaxy spectra down to around seven numbers, while having a clearer picture of what produces the natural dimensionality. The first point to clarify is that in comparison to the first paper, we see that seven rather than five components are required to describe spectra, despite the more appropriate technique. This can easily be explained by the addition of a wider variety of galaxies to the sample, rather than limiting the analysis to ULIRGs.

Having used this more appropriate technique, what do we learn in terms of astrophysics?

Out of the seven components in my final NMF set, two of the components show distinct signatures of AGN activity, while three could be associated with star formation. This confirms AGN activity and star formation activity are two of the main processes producing variations in mid-infrared galaxy spectra. Out of the two AGN components, one shows broad silicate emission associated with the inner region of the torus, while the other contains emission features thought to occur in narrow line regions. I have shown how different mixtures of the two components are required for type 1 and type 2 AGN. This indicates that the viewing angle is another key factor in producing variation of galaxy spectra. The fact that both are required for the majority of the AGN, would also suggest the clumpy torus model rather than

than the smooth torus model is more appropriate for the AGN dusty torus.

We can also learn a great deal about relationships between features by examining how they separate as you vary either the input sample or the allowed number of components. For example, in the simplest NMF sets, PAH emission appears alongside colder dust emission at longer wavelengths. As the number of allowed components is increased, the colder dust slope separates out into other components. This suggests that although there is some correlation between the colder dust and PAH emission, it is not as strong as the correlation between the PAH features. By limiting the sample to PAH dominated objects and re-running the NMF analysis, the algorithm no longer has to contend with AGN emission, the additional freedom can be used to separate out PAH features to components that share similarities to a neutral and ionised PAH component found in resolved spectro-imagery data of galactic PDRs.

### 6.1.1 Future work

The derived NMF components have the potential to be used for further studies into galaxy evolution. One avenue already being explored is using the NMF components to separate out the contribution from AGN and star formation to the infrared luminosity function. Wu et al. (2011) carried out a similar analysis using the IRS spectra from specific objects as templates for star formation and AGN, and then separating out the emission in the 5MUSES sample, a mid-infrared spectroscopic survey of 330 galaxies with 24 $\mu$m flux densities 5 mJy $< f(24\mu\text{m}) <$ 100mJy. By using the NMF components, we are using spectral components that are linked to the actual physical environments such as AGN and star formation unlike templates based on specific objects.

The components could also be used to study the connection between AGN and star formation by looking at the correlation between NMF contributions and data from other wavelengths such as X-ray or radio.

A far more ambitious route for this type of analysis would be to extend the data to other wavelengths, such as optical spectra. By combining optical and infrared data, non-linear physical processes such as dust extinction and absorption will play a far greater role. Any multivariate based analysis approach will have to account for this and so simple linear type approaches such as PCA and NMF will no longer be appropriate. However, non-linear based matrix factorisation techniques

are already being developed and although some of these were explored in paper two and deemed to be inappropriate, other more advanced techniques or a combination of techniques could be used to analyse and extract the information contained within multi-wavelength data.

## 6.2 Radiative transfer of SLEDs

The third paper in this thesis contained work using the non-LTE based RADEX code to constrain the physical parameters of molecular gas from observations of the CO SLED in the nearby spiral galaxy IC342. Unlike other approaches that use a grid based $\chi^2$ approach to fit the CO SLEDs (e.g. Rangwala et al., 2011), I used the nested sampling routine, *Multinest*, discussed in Section 1.4.2 to fully sample the posterior parameter space and calculate the Bayesian evidence in order to compare one and two component model fits.

I first fit a one component model to the CO SLED, but find the lowest CO line is underestimated. A two component model provides a better fit, with one component placing tight constraints on the warm gas, and the second component placing looser constraints on a colder gas component.

My model fit provides density, temperature, column density and gas mass constraints which can be used as a prior for more complicated analysis with PDR and XDR models. The constraints on the warm gas provided by the non-LTE RADEX modelling of IC342, are in agreement with the CO ratios provided by PDR fits, suggesting PDRs are consistent with the observed CO SLED, and hence star formation is the dominant heating source in IC342.

Paper four contained the modelling of $H_2O$ lines from HFLS3, a starburst galaxy at redshift of 6.34, one of the most distant starburst galaxies ever detected. It was shown the assumption of collisional excitation as the main mechanism for excitation is unable to explain observations, even for the lower energy lines. Infrared pumping is suggested as an alternative mechanism, with the infrared radiation coming from star forming regions. This is a good example of how the interstellar medium is home to numerous complex interactions between radiation and matter and that simple models such as RADEX will not always be appropriate.

Indeed, there lies a far greater question of whether the average molecular gas properties of galaxies can be modelled as one or two clumps of gas. By using

RADEX to model the unresolved CO SLEDs from galaxies, we implicitly assume that the summation of CO emission from the different regions within a galaxy, gives rise to an integrated CO SLED whose parameters such as temperature are the average temperature of the different regions. If this is not the case, the constrained parameters from models such as RADEX will have no meaningful physical interpretation. With Herschel and ALMA providing resolved emission for nearby galaxies, we are now in a position to begin examining this question.

### 6.2.1 Future work

My RADEX and *Multinest* modelling setup is currently being used to model the CO SLEDs of ULIRGs observed with Herschel as part of the Herschel ULIRG Survey (HERUS), which combines PACS and SPIRE observations of nearly all ULIRGs with a 60 $\mu m$ flux greater than $\approx$ 1.7Jy. Figure 6.1 shows the one sigma filled contours for temperature and density from the one component RADEX fit to 5 ULIRGs. The 5 CO SLEDs have at least one low CO transition and 3 mid J transitions observed with the SPIRE FTS. There is already significant variation in parameters between 5 objects.

Unlike IC342, many of the objects in HERUS will have a significant contribution from both AGN and star formation. By modelling a large sample of ULIRGs, we will be able to determine if the molecular gas is at similar conditions, whether the conditions vary and if so by how much? We will also be able to investigate whether physical conditions of the molecular gas are related to other global properties such as total infrared luminosity or diagnostics from the mid-infrared (e.g. NMF components).

Examining whether such relations exist, will help us understand whether observations of the molecular gas for higher redshift objects, which will become more accessible with ALMA, can be easily related to star formation or AGN activity, and hence the wider picture of galaxy evolution, or whether the physical and chemical interactions behind the integrated observations of molecular gas are just too complex to interpret in this simple framework.

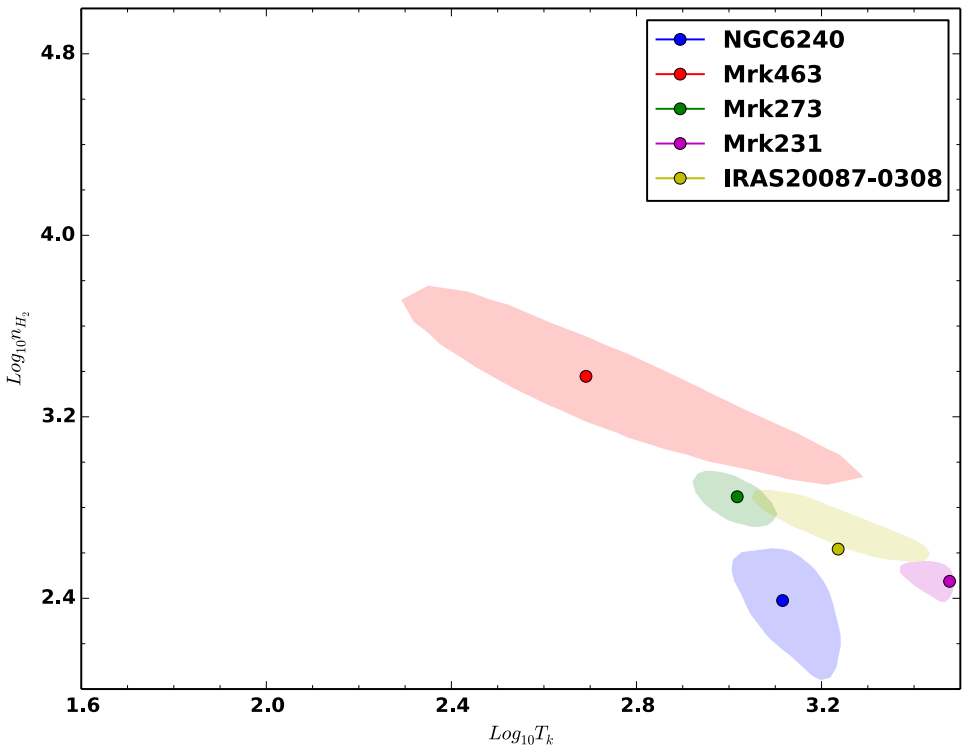


Figure 6.1: The one sigma contours for five ULIRGs, with points indicating maximum likelihood result for each ULIRG.

# Bibliography

Rangwala N., et al. (2011). Observations of arp 220 using herschel-spire: An unprecedented view of the molecular gas in an extreme star formation environment. *The Astrophysical Journal*, 743:94. Cited on 160

Spoon H. W. W., Marshall J. A., Houck J. R., Elitzur M., Hao L., Armus L., Brandl B. R., Charmandaris V. (2007). Mid-infrared galaxy classification based on silicate obscuration and pah equivalent width. *The Astrophysical Journal*, 654:L49. Cited on 157

Wu Y., et al. (2011). The mid-infrared luminosity function at z ¡ 0.3 from 5muses: Understanding the star formation/active galactic nucleus balance from a spectroscopic view. *The Astrophysical Journal*, 734:40. Cited on 159